



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA DE TELECOMUNICACIÓN

INGENIERÍA DE TELECOMUNICACIÓN-LADE

PROYECTO FIN DE CARRERA

DIAGNÓSTICO DE MODELO Y SELECCIÓN DE VARIABLES PARA MÉTODOS DE APRENDIZAJE ESTADÍSTICO APLICADOS A EFECTIVIDAD PROMOCIONAL

Autor: Cristina Soguero Ruiz

Tutor: José Luis Rojo Álvarez

Cotutor: Inmaculada Mora Jiménez

Curso académico 2010/2011

TÍTULO: *Diagnóstico de modelo y selección de variables para métodos de aprendizaje estadístico aplicados a efectividad promocional.*

AUTOR: *CRISTINA SOGUERO RUIZ*

TUTOR: *JOSÉ LUIS ROJO ÁLVAREZ*

COTUTORA: *INMACULADA MORA JIMÉNEZ*

La defensa del presente Proyecto Fin de Carrera fue realizada el día de de 2011, siendo calificada por el siguiente tribunal:

PRESIDENTE:

SECRETARIO:

VOCAL:

Habiendo obtenido la siguiente calificación:

Presidente

Secretario

Vocal

Agradecimientos

La culminación de este proyecto¹ cierra una etapa importante de mi vida, no sólo a nivel académico sino también personal. Han sido 6 años llenos de alegría, de esfuerzo, de metas y de objetivos conseguidos que quisiera agradecer y dedicar a todas las personas que han estado a mi lado y que han hecho posible que esta etapa llegue a su fin.

En primer lugar, quiero dar las gracias a la Universidad Rey Juan Carlos, “mi uni”. Una universidad joven pero con un enorme capital tanto humano como docente preparado, cercano y preocupado de sus alumnos que hará que en poco tiempo se convierta en una Universidad de referencia.

Mi especial agradecimiento a mis tutores José Luis Rojo Álvarez e Inmaculada Mora Jiménez por todo el apoyo y esfuerzo prestado durante la realización de este proyecto, GRACIAS. Siempre les agradeceré la confianza prestada en mí y las oportunidades que me han facilitado.

También me gustaría agradecerle a María Pilar Martínez Ruiz, profesora de la UCLM, la colaboración prestada en el mismo.

A A.T. Kearney, y en especial, a José Ignacio Nieto Puertes y a Carlos Martínez Zarco por la ilusión y la confianza que han puesto en la realización de este proyecto.

A todas las personas que he conocido en Telefónica I+D, y a mis compañeros de despacho del CECL.

A toda mi familia, y en especial, a mis padres y a mi hermana. Es difícil expresar en un párrafo todo lo que les debo, porque es mucho. Darles las gracias por estar a mi lado día a día, por darme fuerza y por ayudarme a conseguir mis objetivos, por escucharme cuando lo he necesitado, por apoyarme, por ser todo. Espero que el tiempo me deje recompensar esa deuda que tengo con ellos. GRACIAS.

A Vanesa, a Rosario, a Abraham por estar a mi lado en estos últimos años de mi carrera, por la el apoyo prestado en los momentos más necesarios, por la amistad que nos ha unido. A María, a Mario, a Mochi, a todos vosotros que habéis significado algo para mí en este periodo. A Elena, a Javi, a todos mis amigos/as que han sabido entender ese repetido no, gracias por los momentos que hemos pasado juntos.

A Juan, por su apoyo incondicional, por entenderme, por escucharme, por ser paciente, sin él, esto no hubiera sido posible.

¹Este proyecto ha sido financiado parcialmente por el proyecto de investigación “Extracción de Características y Selección de Variables en Registros Multivariados Prolongado” (referencia TEC2010-19263).

Ever tried. Ever failed. No matter. Try Again. Fail again. Fail better.

Samuel Beckett

Curts Goetz

La esperanza es el sueño del hombre despierto.

Aristóteles

Resumen

La inestabilidad económica de los últimos años está produciendo una disminución generalizada de las ventas y en concreto de los productos de alimentación. Esta situación ha hecho que muchos distribuidores minoristas hayan puesto en marcha acciones promocionales como el descuento directo y las promociones en cantidad (*3x2*). La información digital disponible hoy en día ha provocado una evolución en el desarrollo de estas actividades, y debido a su potencial, los métodos de aprendizaje estadístico han empezado a cobrar verdadera importancia para conseguir aumentar el volumen de ventas. En el presente proyecto se utilizan los métodos de aprendizaje estadístico para analizar el comportamiento de las promociones en términos de unidades vendidas, realizando una comparación estadística detallada entre diferentes métodos para determinar de forma objetiva cuál de ellos ofrece mejores prestaciones.

La finalidad de este trabajo es proponer un procedimiento operativo para el diagnóstico de modelo y la selección de variables utilizando técnicas estadísticas en aplicaciones de efectividad promocional. En concreto, se han analizado las promociones realizadas por un distribuidor minorista en 6 productos de la categoría de leche y en 14 productos de la categoría de cerveza durante un año, y para ello se han realizado diferentes experimentos. El primero ha consistido en comparar las prestaciones de cuatro métodos de aprendizaje estadístico: k-NN (*k-Nearest Neighbors*), GRNN (*General Regression Neural Network*), MLP (*Multi Layer Perceptron*) y SVM (*Support Vector Machine*) en términos absolutos utilizando el MAE (*Mean Absolute Error*) como figura de mérito. Se ha contrastado si algún método es significativamente mejor que otro mediante un test estadístico no paramétrico, basado en remuestreo *bootstrap*. Esta metodología se ha utilizado posteriormente para contrastar las prestaciones del esquema ν -SVM diseñado con núcleo RBF y con núcleo semiparamétrico, para analizar los elementos de diseño del MLP y comprobar la conveniencia de incluir determinadas variables (en general de naturaleza dicotómica) en los modelos promocionales.

En conclusión, las técnicas de aprendizaje estadístico y la utilización del test *bootstrap* propuesto permiten extraer información relevante en el análisis de la efectividad promocional.

Índice general

1. Introducción	15
1.1. Motivación	15
1.2. Objetivo	16
1.3. Metodología y estructura de la memoria	16
2. Revisión del ámbito promocional	19
2.1. Promoción de ventas	19
2.1.1. Concepto	19
2.1.2. Clasificación	22
2.2. La curva de respuesta promocional	23
2.2.1. Elementos influyentes	24
2.2.2. Tipos de curva	25
2.2.3. Distribuidor minorista	26
2.3. Estudios previos en promociones de ventas	29
3. Métodos de aprendizaje estadístico	35
3.1. Introducción al aprendizaje estadístico	35
3.2. Métodos de regresión basados en aprendizaje estadístico	37
3.2.1. k-NN	37
3.2.2. Redes Neuronales Artificiales. GRNN y MLP	37
3.2.3. SVM	40
3.3. Figuras de mérito y técnicas de evaluación	46
3.3.1. Figuras de mérito	46
3.3.2. Validación cruzada	47
3.3.3. Evaluación con remuestreo <i>bootstrap</i>	48

4. Bases de datos y modelos de respuesta promocional	53
4.1. Bases de datos	53
4.1.1. En la categoría de leche	53
4.1.2. En la categoría de cerveza	54
4.2. Modelos de respuesta promocional	54
4.3. Análisis a nivel de producto	57
5. Análisis y resultados con datos minoristas	71
5.1. Introducción	71
5.2. Elección de parámetros libres	72
5.3. Experimento 1: comparativa de métodos	76
5.4. Experimento 2: comparativa mediante <i>bootstrap</i> pareado	78
5.4.1. Leche	79
5.4.2. Cerveza	81
5.4.3. Conclusiones	82
5.5. Experimento 3: ν -SVM semiparamétrica	84
5.6. Experimento 4: MLP, salida múltiple y salidas individuales	91
5.6.1. Salida múltiple vs salidas individuales	91
5.6.2. Función de activación	100
5.7. Experimento 5: selección de variables	101
6. Conclusiones y líneas futuras	111
6.1. Conclusiones	111
6.2. Líneas futuras	112
APÉNDICES	114
A. Comparativa de figuras de mérito	115
B. Presupuesto del proyecto	123
B.1. Coste de los medios materiales	123
B.2. Coste del personal	123
B.3. Coste de dirección	124
B.4. Coste total	124

Acrónimos y abreviaturas

ANN	<i>Artificial Neural Network</i>
GRNN	<i>General Regression Neural Network</i>
k-NN	<i>k-Nearest Neighbors</i>
LOO	<i>Leave-one-out</i>
MAE	<i>Mean Absolute Error</i>
MSE	<i>Mean Square Error</i>
MLP	<i>Multilayer Perceptron</i>
MMDD	<i>Marca del distribuidor</i>
RBF	<i>Radial Basis Function</i>
RMAE	<i>Root Mean Absolute Error</i>
RRMSE	<i>Root Relative Mean Square Error</i>
ROI	<i>Return On Investment</i>
SVM	<i>Support Vector Machine</i>

Introducción

1.1. Motivación

En la coyuntura económica actual, caracterizada por una época de inestabilidad económica, y ante la alteración de las pautas tradicionales de compra de los consumidores en España, los distribuidores minoristas de productos de alimentación se han visto obligados a modificar las decisiones comerciales que habitualmente implementaban en sus tiendas [1] [2]. Un ejemplo de esta situación es la reducción de la venta de productos de alimentación que se ha producido recientemente en España, de modo que muchos distribuidores minoristas han decidido implementar medidas para incrementar las ventas. Entre estas acciones cabe mencionar un uso más elevado de la marca de distribuidor [3], y la realización de más actividades de promoción de ventas, entre las que destacan los descuentos de precio promocionales, los folletos publicitarios y las promociones en cantidad (como las del tipo *3x2*) [1] [2]. No es de extrañar, por tanto, que la promoción de ventas se haya convertido en los últimos años en una herramienta fundamental en la estrategia de marketing de los distribuidores minoristas, y que la inversión realizada en promociones de ventas haya llegado a representar elevados porcentajes de su inversión anual de marketing, con porcentajes que incluso sobrepasan el 50% [4] [5].

Las implicaciones económicas derivadas de las promociones de ventas, así como la mayor información a la que potencialmente se puede tener acceso para analizar tanto la inversión realizada como los resultados generados, hacen que la caracterización precisa de la promoción de ventas constituya una de las cuestiones de investigación más relevantes que se plantean hoy en día.

Respecto a los modelos de análisis utilizados para explicar los comportamientos de las promociones de ventas, el énfasis ha pasado de modelos teóricos de promoción de ventas a modelos empíricos. Los *modelos teóricos* se fundamentan en la explicación del comportamiento de compra del consumidor, basándose en conceptos psicológicos y sociológicos. Por contra, los *modelos em-*

píricos permiten construir, de forma progresiva, un cuerpo de conocimiento sobre la promoción de ventas a partir de la sucesiva validación empírica de distintos efectos según las diferentes técnicas promocionales, basándose en el análisis de bases de datos históricas.

Es más, en las últimas décadas, y para tratar de entender mejor ciertos comportamientos de las promociones de ventas, se ha producido una evolución del estudio basado en métodos estadísticos clásicos al basado en técnicas de aprendizaje máquina, que optimizan un criterio de desempeño usando datos de ejemplo o la experiencia pasada [6] [7]. El objetivo de los métodos de aprendizaje máquina es encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto, facilitando al ser humano la posibilidad de adquirir nuevos conocimientos sobre el comportamiento del consumidor, mejorar el rendimiento de las operaciones de marketing y estimar la curva de respuesta promocional. Su importancia ha llevado a un conocimiento más profundo en este área en muchos aspectos. Sin embargo, existen comportamientos en promociones de ventas que aún no han sido estudiados con suficiente detalle, y es por ello que surge la necesidad de profundizar en la caracterización de las promociones de ventas usando métodos empíricos [8] [9] [10].

1.2. Objetivo

El objetivo de este trabajo es proponer un procedimiento operativo para el diagnóstico de modelo y selección de variables para métodos de aprendizaje estadístico en aplicaciones de efectividad promocional.

En este proyecto se utiliza un enfoque empírico basado en métodos de aprendizaje máquina para analizar la variación de las ventas en productos pertenecientes a dos categorías (leche y cerveza) como respuesta a los descuentos promocionales en un entorno minorista. Así, los dos pilares del trabajo desarrollado son la promoción de ventas y los métodos de aprendizaje basados en ejemplos. Los datos para realizar este estudio han sido facilitados por A.T. Kearney S.A. y proceden de bases de datos reales de distribuidores minoristas.

1.3. Metodología y estructura de la memoria

La metodología seguida en el presente proyecto consta de las siguientes etapas:

- Revisión bibliográfica detallada del comportamiento de las promociones de ventas, tanto a nivel académico como a nivel empresarial. Con esta búsqueda bibliográfica se pretende conocer la clasificación de las promociones de ventas, la influencia de los descuentos de precios y su análisis en un entorno minorista.
- Presentación de las bases de datos de dos categorías de productos consideradas (leche y

cerveza), así como el comportamiento individual de cada producto frente a las promociones realizadas.

- Comparación de las prestaciones obtenidas al estimar el número de unidades vendidas con distintos métodos de aprendizaje máquina, en concreto, k-NN (del inglés *k-Nearest Neighbors*), GRNN (del inglés, *General Regression Neural Network*), MLP (del inglés *Multi Layer Perceptron*) y SVM (del inglés *Support Vector Machine*).
- Aplicación de inferencia no paramétrica basada en remuestreo *bootstrap* para decidir si un método ofrece prestaciones significativamente superiores a otro. Este remuestreo permite obtener la distribución muestral de los estadísticos de interés [11] proporcionando un test de corte operativo, una descripción detallada de las prestaciones, y una herramienta de apoyo a la toma de decisiones en modelado promocional.

La estructura de la memoria es la siguiente:

- En el Capítulo 2 se presenta una revisión de la literatura sobre promociones de ventas. De una parte, se expone el concepto de promoción de ventas, su clasificación, características y principales objetivos. Se indican los elementos que intervienen en la curva de respuesta promocional y los tipos de curva existentes. De otra, se presenta una revisión de las líneas de investigación relacionadas con acciones promocionales en el área de marketing.
- En el Capítulo 3 se exponen los fundamentos teóricos de aprendizaje máquina necesarios para el desarrollo del presente trabajo. En concreto, se analizan diferentes métodos de regresión no lineal (k-NN, GRNN, MLP y SVM), se presentan diferentes técnicas de evaluación de prestaciones, figuras de mérito usadas para el diagnóstico y selección de variables y, por último, un test no paramétrico basado en remuestreo *bootstrap*, que permite tomar decisiones sobre la conveniencia de los distintos modelos de ventas promocionales con una descripción estadística detallada y operativa.
- En el Capítulo 4 se describen las bases de datos de las categorías de leche y cerveza, así como los modelos de respuesta promocional. Se realiza un análisis promocional a nivel de producto.
- El Capítulo 5 presenta los experimentos desarrollados, que han permitido determinar los parámetros libres de los esquemas de aprendizaje estadístico, realizar mejoras de los modelos de promoción y determinar la conveniencia de la inclusión de la línea de base y/o de las variables dicotómicas más relevantes.
- En el Capítulo 6 se exponen las conclusiones y líneas futuras de investigación.

Revisión del ámbito promocional

Debido a que la promoción de ventas es el objeto de estudio del presente proyecto, se describe en primer lugar qué se entiende y qué engloba dicho concepto. Además, se estudia la influencia que tienen los descuentos de precios aplicados en un producto sobre su volumen de ventas, distinguiendo entre descuentos propios y descuentos del competidor frente al propio. Se lleva a cabo un análisis más detallado de las promociones realizadas por distribuidores minoristas. Por último, se presentan investigaciones previas en promociones de ventas, tanto a nivel académico como empresarial.

2.1. Promoción de ventas

2.1.1. Concepto

Existen muchas definiciones de promoción de ventas [12]. Sin embargo, según [4] no existe un concepto generalmente aceptado por la totalidad de los autores. No obstante, todas ellas coinciden en que es un incentivo promocional durante un periodo de tiempo determinado [5]. Entre estas definiciones, se encuentra la adoptada por la American Marketing Association: *conjunto de actividades de marketing dirigidas a través de medios de comunicación masivos o no masivos, orientadas a ejercer presión durante un periodo de tiempo determinado y limitado, a nivel del consumidor, del minorista o del mayorista, con el fin de estimular la elección, aumentar la demanda del consumidor final o mejorar la disponibilidad de los productos.*

Algunas investigaciones [13] consideran que la promoción de ventas no sólo es un elemento de marketing, sino que se engloba dentro de la actividad estratégica de la empresa. Por tanto, *la orientación al mercado supone una reconciliación e integración de dos aspectos. El carácter integrado de las actividades de marketing en la organización, superando su aislamiento funcional, y la integración de la visión estratégica, mediante el reconocimiento de la importancia del entorno*

[13]. Con este enfoque aparecen nuevas definiciones del concepto de promoción de ventas, como la siguiente: *conjunto de actividades realizadas por la empresa, como consecuencia del ofrecimiento de un incentivo, durante un periodo determinado de tiempo, que implican la coordinación de diferentes áreas funcionales de la empresa, para la consecución de determinados objetivos comerciales* [5].

Como características relevantes de la promoción de ventas se pueden destacar las siguientes [14] [15]:

- **Selectividad.** La promoción de ventas suele tener límites y objetivos muy claros. Entre ellos se puede encontrar el incremento de la demanda de un producto en particular, el aumento de la presencia de una marca en determinados tipos de establecimiento, la obtención de resultados en periodos de tiempo específicos, el apoyo a la publicidad o a la venta personal de un producto en una determinada región.
- **Intensidad y duración.** La efectividad de la promoción de ventas se pone de manifiesto cuando se implementa de forma intensa y durante un corto periodo de tiempo.
- **Resultados a corto plazo.** La promoción de ventas se caracteriza por proporcionar resultados beneficiosos inmediatos, tanto para el vendedor como para el consumidor. Al vendedor no sólo le permite autofinanciarse, sino también aumentar su rentabilidad. Al consumidor le permite aprovecharse de descuentos directos sobre el precio del producto, adquirir una cantidad adicional del mismo o conseguir productos diferentes al del objeto de la compra.
- **Imagen de producto o de empresa.** La promoción influye tanto en la imagen del producto como de la empresa, por lo que debe estar correctamente integrada en la estrategia de marketing de la misma. Así, la promoción de ventas deber ser incorporada en el plan estratégico de marketing, y esto implica establecer los objetivos de la promoción, que tienen que ser claramente fijados porque van a influir en el comportamiento de compra del consumidor.

Del concepto de promoción de ventas se pueden extraer tres objetivos fundamentales [16]:

1. Incrementar el nivel de atención de los consumidores, proporcionando información relevante que capte su atención.
2. Mejorar el desempeño de marketing de los intermediarios y de los vendedores [17].
3. Completar y coordinar las actividades de publicidad, fuerza de ventas y relaciones públicas.



Figura 2.1: *Principales objetivos de la promoción de ventas.*

En líneas generales, las campañas de promoción tienen cuatro grandes objetivos, los cuales se indican en la Figura 2.1 [4] :

1. **Información sobre la empresa o sobre los productos.** En este sentido, la promoción de ventas hace referencia a la acción de comunicación promocional dirigida a potenciar la apertura de mercados a todos los niveles: red de ventas, canales de distribución y consumidores. Todo ello no sólo para conseguir un primer lanzamiento, sino también con objeto de alcanzar la aceptación de la novedad de manera reiterada y repetitiva.
2. **Introducción de productos o de la empresa.** Consiste en el apoyo promocional que aporta para lograr que el público tenga conocimiento del producto o del mercado de un modo económico, tanto en tiempo como en dinero.
3. **Motivación e imagen.** Aparece cuando se desea crear o reforzar acciones positivas hacia el producto o la empresa de una forma eficaz.
4. **Animación de las ventas.** La promoción de ventas resulta de gran utilidad para la consecución de incrementos en la venta del producto, y por tanto, su expansión en el mercado. Este objetivo abarca la solución de problemas importantes para la empresa, como pueden ser: subidas imprevistas de precios, niveles bajos de ventas, baja rotación de stocks o escasa fidelidad a la marca. El coste de la promoción se autofinancia por el incremento del margen que genera.



Figura 2.2: Clasificación de las promociones de ventas según el tipo de campaña promocional.



Figura 2.3: Clasificación de las promociones de ventas en función del iniciador y destinatario de la misma.

2.1.2. Clasificación

Existen diversos elementos a tener en cuenta cuando se realiza una clasificación de la promoción de ventas. En concreto, se suelen distinguir dos clasificaciones atendiendo a los siguientes criterios:

- Según el tipo de campaña promocional [18], éstas se clasifican por: duración, objetivos, resultados y forma de la promoción (véase la Figura 2.2). En este sentido, se puede tener, por ejemplo, una promoción permanente cuyo objetivo sea introducir un nuevo producto para alcanzar nuevos segmentos de mercado.
- Según los iniciadores y destinatarios [4] de la promoción de ventas, éstas se clasifican en

Agente	Objetivos
Comercial	<ul style="list-style-type: none"> • Motivar a los distribuidores para que adquieran las mercancías. • Aumentar y mantener la presencia en el canal de distribución. • Utilizar las promociones como una herramienta competitiva. • Incrementar la eficacia del trabajo realizado por los distribuidores. • Favorecer la integración del producto en la red de comercialización. • Fidelizar a los distribuidores actuales de la empresa actuales de la empresa. • Distribuir la marca a través de nuevos intermediarios, en nuevas áreas geográfica y formatos de establecimiento.
Del fabricante al consumidor final	<ul style="list-style-type: none"> • Conseguir la prueba del producto, acercando el producto al consumidor. • Favorecer la introducción de nuevos productos. • Atraer a los nuevos compradores. • Crear y mantener el hábito de compra hacia el producto. • Favorecer la acumulación del producto por parte del consumidor. • Reforzar la publicidad e imagen de marca del producto. • Capitalizar ciertas zonas, temporadas o acontecimientos, intentando atraer nuevos compradores • Incentivar a los consumidores para que adquieran marcas de mayor prestigio. • Neutralizar la publicidad o la promoción de ventas de los competidores.
Minorista	<ul style="list-style-type: none"> • Incrementar las ventas de la categoría de producto. • Aumentar el tráfico y el número de clientes al establecimiento. Aumentar la frecuencia de compra • Ganar y mantener la lealtad de los clientes. • Estimular la demanda del producto y aumentar el volumen de ventas. • Reforzar la imagen del establecimiento. Crear una imagen de competitividad en precios. • Disminuir los periodos de estacionalidades en las ventas. • Aumentar las ventas de productos con marcas de distribuidor.

Figura 2.4: *Objetivos de las promociones de ventas en función de los agentes que la realizan (tomado de [4]).*

promociones comerciales, promociones al consumidor y promociones minoristas. La Figura 2.3 muestra la relación existente entre el tipo de promoción, el iniciador y el destinatario de la misma.

Como se muestra en la Tabla 2.4, los objetivos de la promoción de ventas cambian en función del agente que realiza la promoción de ventas.

2.2. La curva de respuesta promocional

La curva de respuesta promocional constituye un tema fundamental en la literatura de marketing. Representa el volumen de ventas de un producto en función de los descuentos de precios aplicados en el mismo. A lo largo del presente proyecto se analizan dos tipos de efectos. El prime-

ro, conocido como efecto propio del descuento, indica cómo responden las ventas de un producto a los descuentos temporales en el precio del propio producto. El segundo, que se conoce como efecto cruzado del descuento, se refiere a cómo responden las ventas de un producto frente al descuento aplicado en otro producto.

2.2.1. Elementos influyentes

Uno de los principales objetivos cuando se realiza una promoción es el incremento de las ventas. El cumplimiento de dicho objetivo depende de tres elementos, a saber: las características del descuento del propio precio, los descuentos de precio del competidor frente al descuento propio, y las características de la promoción. A continuación se describen los efectos que responden a cada uno de estos elementos.

Efectos del descuento propio

- *Efecto de intervalo.* Es el mínimo descuento que se debe aplicar sobre el precio de venta de una determinada referencia (marca, producto, etc.) para que, cuando ésta se ponga en promoción, se produzca un incremento de las ventas [12].
- *Efecto de saturación.* Es aquel nivel de descuento a partir del cual las ventas promocionales no se incrementan, sino que se estancan, pudiendo incluso descender. Existen dos motivos principales por los que este efecto puede ocurrir. El primero hace referencia a la cantidad máxima que los consumidores pueden almacenar y/o consumir si los productos sobre los que se aplica la promoción son perecederos [10]. El segundo, que se conoce como *efecto del descuento descontado*, se relaciona con la mente del consumidor, de modo que la percepción de los descuentos resulta menor que los descuentos reales ofrecidos [12].

Efectos del descuento del competidor frente al propio

- *Efectos cruzados de las promociones sobre las ventas de otras marcas o categorías de producto.* Los descuentos aplicados en determinados productos dentro de una marca o categoría pueden influir en las ventas obtenidas en otras marcas o categorías de producto. Se ha demostrado que los resultados obtenidos son asimétricos en función de la marca que se ponga en promoción, es decir, considerando la misma categoría de producto, si se pone en promoción una marca de precio alto se produce un incremento en las ventas considerablemente mayor que el generado cuando se pone en promoción una marca de precios bajos [19].
- *Efectos de interacción entre promociones simultáneas dentro de la categoría.* Los descuentos realizados de forma simultánea en varias alternativas sustitutivas pueden tener repercusio-

nes, tanto en sus propias ventas como en las ventas del resto de marcas competidoras [20].

Efectos que responden a las características de la promoción

- *Efectos de interacción entre diferentes actividades y soportes promocionales.* Existen diferentes actividades y soportes promocionales, y dependiendo de cuál y de cuántos soportes o actividades se usen, la forma de la curva de respuesta promocional será distinta. En concreto, la combinación del folleto y la exposición especial es el tipo de promoción que consigue un mayor efecto. En segundo lugar se encuentra la exposición especial, seguido del descuento temporal de precio, y por último, los folletos [21]. Se observa que es posible incrementar las ventas en ausencia de descuentos de precio a través de publicidad promocional o con la disposición de los productos en el establecimiento.

2.2.2. Tipos de curva

En función de la existencia de los efectos anteriores, la curva de respuesta promocional puede presentar una forma u otra. El estudio de esta curva permite tomar importantes decisiones sobre la aplicación del descuento promocional y la importancia del mismo. Los principales tipos de curva de respuesta promocional son los siguientes:

- **Forma lineal.** Puede ocurrir que no exista variación en las ventas con el nivel de descuento aplicado, en cuyo caso la pendiente de la recta es nula. Si existe una relación directa entre el nivel de descuento aplicado y el incremento de las ventas, la pendiente es positiva; si la relación es inversa, la pendiente es negativa.
- **Forma de S.** Este tipo de curva presenta dos efectos (de intervalo y de saturación) atendiendo al tipo de descuento propio. A partir de un cierto nivel de descuento se produce un incremento de las ventas (efecto de intervalo), pudiendo ocasionar un estancamiento de las ventas si se supera un cierto nivel de descuento (efecto de saturación) [22].
- **Forma de L.** Este tipo de curva de respuesta promocional sólo presenta efectos de intervalo, pero no de saturación.
- **Forma convexa.** Si la curva de respuesta promocional tiene forma convexa aportará mayores beneficios que si la curva es cóncava [22]: para un mismo descuento de precio, el número de unidades vendidas es mayor si la curva de respuesta promocional es convexa (frente a cóncava).

2.2.3. Distribuidor minorista

El distribuidor minorista es la empresa comercial que vende productos al consumidor final. Son el último eslabón del canal de distribución, y por tanto, el que está en contacto con el mercado. Son importantes porque pueden alterar, frenando o potenciando, las acciones de marketing y merchandising de los fabricantes y mayoristas, influyendo en las ventas y resultados finales de los artículos que comercializan. El comercio minorista compra productos en grandes cantidades a fabricantes o importadores, bien directamente o a través de un mayorista. Sin embargo, vende unidades individuales o pequeñas cantidades al público en general, normalmente, en un espacio físico denominado tienda.

Una vez que el distribuidor minorista ha llevado a cabo las actividades de aprovisionamiento, realiza actividades de marketing para decidir la mejor forma para llevar a cabo la venta de sus productos. Entre los objetivos principales del minorista se encuentran [4]:

- Incrementar las ventas de la categoría de producto.
- Aumentar las ventas de productos con marcas de distribuidor.
- Estimular la demanda del producto y aumentar el volumen de ventas.
- Crear una imagen de competitividad en precios.
- Disminuir los periodos de estacionalidades en las ventas.
- Diferenciarse de los competidores.
- Aumentar el tráfico y el número de clientes en el establecimiento.
- Reforzar la imagen del establecimiento.
- Ganar y mantener la lealtad de los clientes.
- Aumentar la frecuencia de compra.

Se distinguen tres categorías principales de efectos que los distribuidores minoristas pretenden alcanzar en función de los anteriores objetivos [4]:

1. **Efectos de atracción.** Hace referencia a los efectos sobre el tráfico de clientes que acude al punto de venta. A los distribuidores les interesa que un número elevado de clientes visite la tienda.

2. **Efectos de conversión.** Son efectos relativos a las decisiones que los consumidores llevan a cabo en el interior de la tienda para la compra de un producto determinado. Los distribuidores minoristas están interesados en que esas decisiones se materialicen en la compra del producto.
3. **Efectos de gasto.** Se analiza el tamaño de la compra. A los minoristas les interesa que los consumidores finales realicen compras de elevado tamaño.

En la Figura 2.5 se muestran los objetivos del minorista en función de los efectos deseados. El objetivo final es aumentar la rentabilidad obtenida debido a un aumento de las ventas. En la Figura 2.6 se muestran las causas que producen un aumento de las ventas.

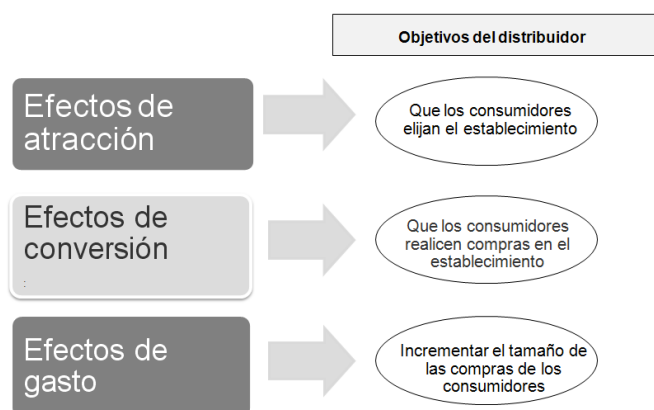


Figura 2.5: *Objetivos del distribuidor minorista en función de los efectos deseados.*



Figura 2.6: *Causas del incremento de ventas del minorista.*

Para cuantificar los efectos de canibalización que la promoción de ventas es susceptible de crear, los autores de [23] definieron en 1993 el modelo descrito en la Figura 2.7. Este modelo considera que sólo existen dos marcas dentro de la categoría, A y B, y que durante el período considerado la marca A se encuentra en promoción. En dicho modelo se tiene que $(m_A + d - r)b_A\eta$ es el beneficio de las ventas adicionales de A; $\pi_c b_A(\eta_c + \eta_t)$ constituye el beneficio incremental para las categorías de productos complementarias vendidas en dicho establecimiento; $\pi_s b_A \eta_t$ representa los beneficios adicionales obtenidos por el tráfico adicional de clientes debido a la promoción de la marca A; $m_B(b_A \cdot \eta_s)$ representa el descenso de beneficio producido al descender las ventas de marcas competidoras dentro de la misma categoría de productos (la marca B, en este caso); y A representa el coste de la promoción de la marca A.

El porcentaje de las ventas promocionales varía en función del establecimiento (Tabla 2.1) y de la categoría de producto (Tabla 2.2). Obsérvese cómo el porcentaje de ventas en promoción, tanto para supermercado como para hipermercado, es elevado. De los sectores mostrados en la Tabla 2.2, los productos de bebida representan el primer lugar en términos de porcentaje de ventas promocionales, tanto en supermercados como en hipermercados.

<p>b_A = ventas semanales minoristas (en unidades) para la marca A. m_A = margen normal minorista por unidad para la marca A. m_B = margen normal minorista por unidad para la marca B. d = descuento del fabricante al minorista para la marca A. r = reducción del precio minorista de la marca A debido a la promoción de dicha marca. η = incremento de las ventas minoristas debido a la promoción de la marca A (% de b_A). η_s = incremento de las ventas de A debido al cambio de marca (% de b_A). η_c = incremento de las ventas de A debido al incremento de las ventas de la categoría de productos de A (% de b). η_t = Incremento de las ventas de A debido al cambio de establecimiento (% de b_A), Por tanto, $\eta = \eta_s + \eta_c + \eta_t$; π_c = Beneficio incremental del establecimiento debido al incremento de las ventas de productos complementarios (por unidad adicional vendida de la categoría de producto). π_s = Beneficio incremental del establecimiento debido al cambio de establecimientos (por ventas adicionales obtenidas por el cambio de establecimiento). A = Coste de promocionar la marca A. Π = Rentabilidad de la promoción minorista.</p>
$\Pi = (m_A + d - r) b_A \cdot \eta + \pi_c b_A (\eta_c + \eta_t) + \pi_s b_A \eta_t - m_B (b_A \cdot \eta_s) - A$

Figura 2.7: Modelo para calcular la rentabilidad de la promoción minorista.

Actividad promocional por canales de distribución	% de las ventas promocionales
Hipermercado	38
Supermercados (1000-2499 m ²)	15,2
Supermercados (400-999 m ²)	20,2
Supermercados (100-399 m ²)	23,1

Tabla 2.1: *Actividad promocional según el formato del establecimiento.*

% Ventas en promoción por sector	% de las ventas promocionales en supermercado	% de las ventas promocionales en hipermercado
Alimentación/Bebidas	14,1	28,1
Productos frescos	13,8	29,6
Droguería/Perfumería	14,9	34,9
Bebé	23,4	35,3

Tabla 2.2: *Porcentaje de ventas en promoción para supermercados e hipermercados.*

2.3. Estudios previos en promociones de ventas

A lo largo de los últimos años, han sido numerosos los estudios relacionados con promociones de ventas. Los principales elementos estudiados están relacionados con las ventas promocionales, las características del descuento, el horizonte temporal (largo y corto plazo), la marca y los efectos psicológicos de los precios promocionales. Algunos de los elementos más analizados han sido [4]:

- Fuentes del incremento de ventas promocionales.
- Eficacia del descuento respecto a la marca propia, otras alternativas sustitutivas, y su interacción con otras actividades promocionales.
- Impacto de la magnitud, profundidad y frecuencia del descuento.
- Efectos a corto plazo de los descuentos de precios promocionales.
- Efectos a largo plazo de los descuentos de precios promocionales.
- Efectos sobre el comportamiento de compra del consumidor en las promociones.

En la Figura 2.8 se observa, dentro de cada uno de los apartados anteriormente enumerados, cuáles son los factores más estudiados en la literatura.

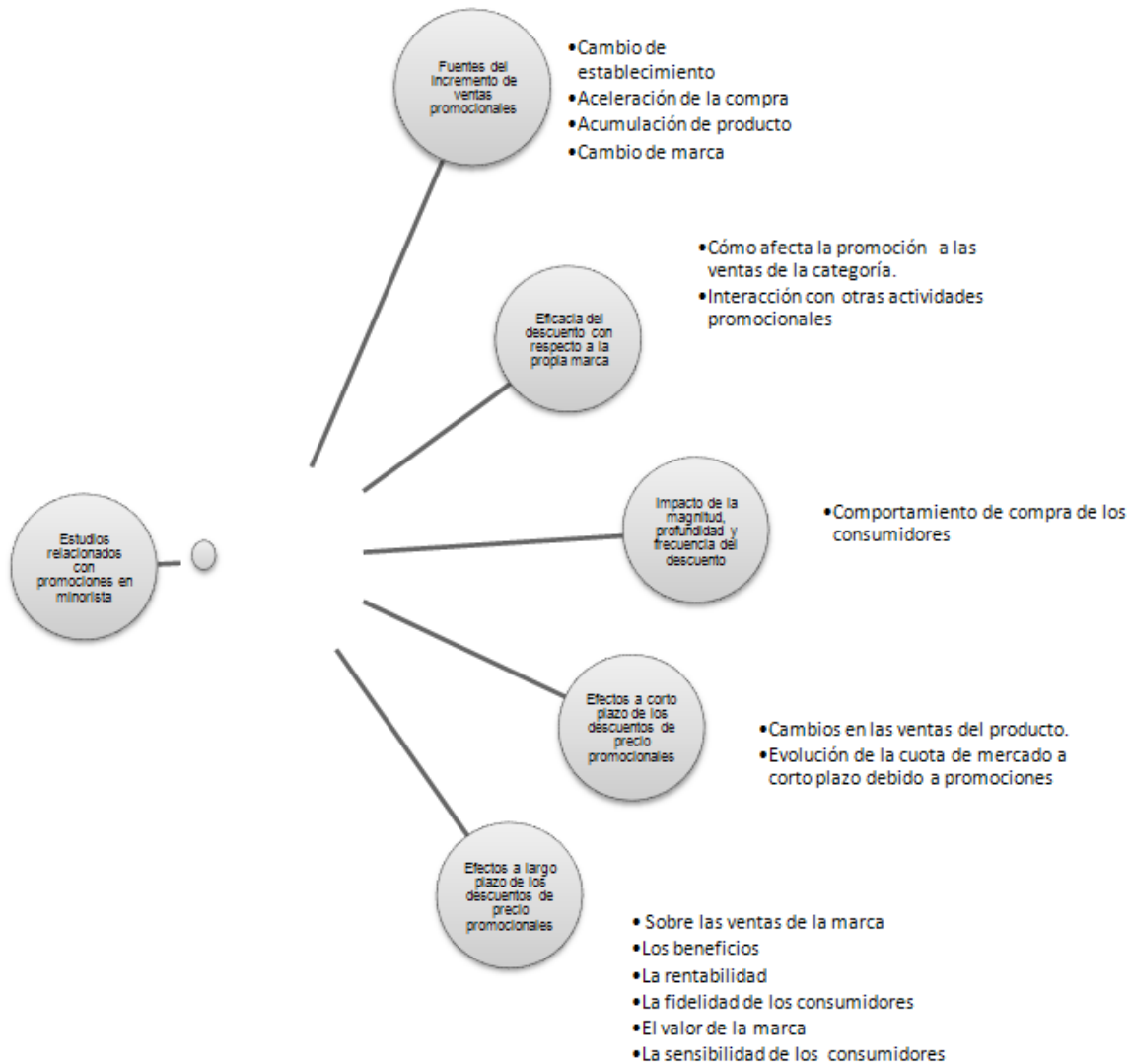


Figura 2.8: Principales factores analizados en promociones de ventas.

A nivel académico

En el ámbito académico se ha incrementado notablemente la investigación en promociones de ventas, tanto en términos cuantitativos como cualitativos. Existen diferentes líneas de investigación que centran su interés en:

- Conocer el efecto de las actividades de promoción de ventas sobre distintos ratios de negocio de los distribuidores minoristas, como las ventas o los beneficios.
- Encontrar cuáles son las características más comunes de los consumidores más propensos al uso de promociones de ventas. Por ejemplo, gran parte de la literatura existente en torno a la primera de las líneas de investigación apuntadas se ha centrado en analizar el impacto de las promociones de precio sobre las ventas de los productos y categorías promocionadas, así como sobre las ventas de otros productos y marcas no promocionadas [8] [10] [19] [20] [24] - [25].
- Investigar el efecto de las promociones de precio sobre la rentabilidad del minorista, encontrando resultados diversos. Por ejemplo, mientras que algunos estudios han detectado cómo las promociones de precio no ejercen un impacto importante sobre los beneficios agregados del establecimiento [26], otros autores ponen de relieve cómo las promociones en productos de gran prestigio o marcas líderes, incrementan el beneficio neto del establecimiento (a pesar de que la magnitud del efecto pueda diverger entre las categorías de productos consideradas). Otras investigaciones llegan incluso a poner de manifiesto la existencia de una relación negativa entre las promociones de precio y la rentabilidad del minorista [26] [27] [28], o a no establecer una relación clara entre estas variables al detectar una influencia positiva, negativa o nula dependiendo de la categoría de producto considerada [29].

Finalmente, sobre la tercera de las líneas de investigación previamente mencionadas, cabe destacar cómo algunas investigaciones han tratado de caracterizar cuáles son los consumidores más propensos al uso de una de las técnicas más utilizadas por los minoristas como son los folletos publicitarios. Estas investigaciones que habitualmente se han llevado a cabo analizando los efectos del uso combinado de los descuentos de precio promocionales y los folletos publicitarios [30] [31] [32], o incluso los folletos publicitarios únicamente [33] - [34], que llegan a desarrollar un análisis detallado de la influencia que los aspectos relacionados con la estructura y composición de los folletos ejercen sobre las ventas del comerciante detallista y el tráfico de clientes que se genera dentro del establecimiento. Por otra parte, [33] [35] [34] tratan de analizar la actitud del consumidor hacia el uso de folletos publicitarios, siendo capaces de entrever algunos de los motivos que les llevan a utilizarlos.

Las principales ideas que se pueden extraer de esta revisión son [4]:

- Cuando se realiza una promoción, en la mayoría de ocasiones hay un incremento de las ventas.
- Las marcas de mayor cuota de mercado presentan menor elasticidad promocional.
- Los efectos cruzados promocionales son asimétricos. La promoción de una marca de precios más elevados repercute negativamente sobre las marcas de precios más reducidos. Un efecto similar ocurre cuando se ponen en promoción diferentes formatos de tamaño de la misma marca.
- La percepción del consumidor sobre una determinada marca influye en la eficacia de la promoción.
- La información que el consumidor perciba de la promoción es un factor muy relevante a tener en cuenta.
- El uso de folletos publicitarios contribuye a incrementar el tráfico de clientes al establecimiento.
- El uso de publicidad promocional y displays influye en las ventas.
- Cuanto mayor es la frecuencia de la promoción, menor es la magnitud del incremento de las ventas.
- La frecuencia promocional provoca cambios en el precio de referencia del consumidor.
- La promoción de ventas en un determinado producto influye sobre las ventas de categorías complementarias y competidoras.

A nivel empresarial

Se realiza una revisión de aquellas empresas que realizan estudios de promociones en marketing. El motivo principal es la importancia de la promoción de ventas como una de las más poderosas armas de la estrategia de marketing para muchas empresas y diferentes mercados. En las últimas décadas, la inversión realizada en promoción de ventas representa un porcentaje muy importante del presupuesto anual de marketing, ya que se ha mostrado como un elemento determinante en el incremento de las ventas. En los últimos años, se ha producido un desarrollo tecnológico que trata de mejorar las herramientas para facilitar la toma de decisiones en promociones de ventas [9]. A continuación se citan algunas empresas que llevan a cabo dicho análisis.

SymphonyIRI Group es una consultora de marketing que ofrece una combinación de investigación de mercado, estudios analíticos, consultoría y aplicaciones tecnológicas. Para enfrentarse a la transformación del mercado de Gran Consumo, SymphonyIRI Group ofrece una

combinación de herramientas de análisis, soluciones en tiempo real y tecnología. En la solución que el grupo denomina Brand Execution, en concreto en servicios de consultoría, se llevan a cabo estudios relacionados con precios y promociones. Aparecen los siguientes productos (<http://www.symphonyiri.com>):

- Price & Promotion Simulator. Es un simulador que, a través de resultados predictivos, ayuda a tomar decisiones estratégicas y de promoción para evaluar los diferentes escenarios de las acciones promocionales y de precios.
- Promo Drivers. Es un modelo que permite analizar la respuesta de un producto a las políticas de precios y promociones. A diferencia de otros análisis, su metodología combina varias fuentes de información.
- Price Drivers. Es una solución que utiliza técnicas estadísticas para cuantificar la sensibilidad al precio y es un soporte para la toma de decisiones.

Martin Long, el director de negocios de la consultora IRI, escribió en 2008 un artículo titulado *Promotional effectiveness -The changing dynamics of marketing promotions in FMCG*, en el cual se analiza la necesidad de cambiar el tipo de promociones realizadas hasta la fecha, argumentando que no son lo suficientemente efectivas. Los fabricantes invierten en promociones entre el 15 % y el 25 % del total del presupuesto destinado a marketing. Se intenta pasar de un escenario en el que las decisiones sobre promociones se tomaban teniendo en cuenta los conocimientos del mercado y la propia intuición, a una situación en la que se pretende construir modelos de demanda que permitan extraer resultados más detallados. Este grado de detalle se consigue gracias al uso de códigos de barras, que permite recopilar una gran cantidad de información sobre compras y usuarios. Con esto, se extrae información sobre canibalización o desplazamiento de los competidores, entre otros. Resalta la importancia obtenida en términos de incremento de ventas. No obstante, como cabe esperar, no se indican las técnicas de análisis aplicadas para conseguir los objetivos mencionados.

Analytic Partners es otra consultora localizada en Norte América, Europa y Asia que presenta la efectividad promocional como uno de sus servicios, junto a la efectividad publicitaria o la predicción de modelos, entre otros. En cuanto a la efectividad promocional, evalúa las actividades promocionales y proporciona la ayuda necesaria para obtener los mayores beneficios posibles. Dicha consultora valora la aplicación de actividades promocionales como cupones, paquetes especiales, descuentos de precios, muestras, *2x1* o, por ejemplo, localizaciones especiales, para tratar de mejorarlas y conseguir mayores ingresos o aumentar la ROI (del inglés *Return On Investment*).

Marketing Analytistics, fundada en 1991, realiza análisis de promociones teniendo en cuenta cuáles son los puntos clave de la variable precio, la sensibilidad al cambiar el precio o el precio

fijado por la competencia. Además, considera variables macroeconómicas, estacionales o relativas a los competidores. En función de la promoción aplicada, evalúa los impulsos de compra y la ROI. La metodología Store Group ha sido testada en estudios académicos independientes. Un grupo de investigadores de las universidades de Cornell, Northwestern, Duke y de la empresa AC Nielsen, encontraron que dicha metodología era el método de estudio más eficaz para el modelado de datos agregados.

Accuris, localizada en Londres, ofrece soluciones para mejorar la efectividad promocional. Pretende aumentar la ROI, hacer que las promociones tengan una mayor contribución tanto en la marca como en la categoría objetivo, o reducir el tiempo de mano de obra para dirigir las actividades promocionales. Implementa un sistema que permite medir la efectividad promocional antes, durante y después de la promoción. Muestra el nivel de impacto de las promociones, tanto en la marca como en la categoría, el efecto canibalización y el desplazamiento de los competidores.

Como conclusión final, en base a la bibliografía revisada y los desarrollos realizados a nivel empresarial, se puede decir que la efectividad promocional está cobrando verdadera importancia en los últimos años, y por ello es objeto de estudio en el presente proyecto. Es importante resaltar que, a diferencia de los estudios realizados en un ámbito académico o de investigación, en los estudios que se realizan a nivel empresarial no se revela la metodología usada para explicar el comportamiento de las promociones de ventas.

Métodos de aprendizaje estadístico

En este capítulo se presentan los conceptos teóricos del aprendizaje estadístico que sustentan el desarrollo presentado en este trabajo. En primer lugar, se explica el concepto de aprendizaje estadístico. En la Sección 3.2 se presentan los cuatro métodos usados (k-NN, GRNN, MLP y SVM) y en la Sección 3.3 se presentan las figuras de mérito y las técnicas de evaluación de las prestaciones de los modelos.

3.1. Introducción al aprendizaje estadístico

En las últimas décadas, debido a la gran cantidad de datos disponibles, ha aumentado el estudio de las técnicas de aprendizaje estadístico para explicar el comportamiento de las promociones de ventas [21]. Estas técnicas permiten generar o diseñar un modelo de regresión $y = f(\mathbf{x})$ que aprenda las relaciones existentes entre un conjunto finito de N ejemplos, cada uno de ellos formado por unas variables observadas (patrón de entrada \mathbf{x}) y su correspondiente salida y . Se desea que el modelo construido pueda generalizar, es decir, obtener una salida adecuada ante patrones de entrada no considerados durante la fase del diseño.

En el caso considerado en este proyecto, la entrada del modelo $f(\cdot)$ corresponde a un vector \mathbf{x} que contiene variables que representan cambios en los precios y variables relativas a las características de la promoción. La salida y corresponde al número de unidades vendidas de un producto determinado. En el ámbito promocional se han usado tres métodos diferentes para estimar el modelo f : métodos paramétricos, métodos no paramétricos y métodos semiparamétricos.

Los métodos paramétricos asumen un modelo funcional cuyos parámetros hay que determinar en función de los ejemplos disponibles. Estos métodos de regresión son los que primero se aplicaron en el ámbito promocional, fundamentalmente por su sencillez y facilidad de interpretación. Entre estas técnicas se encuentra el método de Mínimos Cuadrados Ordinarios [36].

Frente a los métodos paramétricos, los métodos no paramétricos no asumen ninguna forma funcional para el modelo $f(\cdot)$. El estimador de *Nadaraya-Watson*, basado en una función núcleo $\kappa(\cdot)$, es uno de los estimadores no paramétricos más utilizados [37]:

$$f(\mathbf{x}) = \frac{\sum_{i=1}^N \kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{\sigma}\right) y_i}{\sum_{i=1}^N \kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{\sigma}\right)} \quad (3.1)$$

Puesto que el estimador hereda las características de la función núcleo, una elección común de ésta es la función gaussiana, por ser suave, derivable y decrecer monótonamente a medida que se aleja de su centro:

$$\kappa\left(\frac{\mathbf{x} - \mathbf{x}_i}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}} \quad (3.2)$$

donde σ , que corresponde a la anchura de la función núcleo, es un parámetro de suavizado. Valores de σ excesivamente altos conducen a estimaciones demasiado suaves, mientras que valores de σ demasiado pequeños conducen a modelos muy sensibles al ruido [38]. La elección del valor de σ se suele realizar minimizando alguna función de coste.

La principal ventaja de los métodos no paramétricos es su flexibilidad; sin embargo, si se comparan con los métodos paramétricos, tienen menor velocidad de convergencia y requieren un mayor número de ejemplos. Como alternativa se ha propuesto el uso de métodos semiparamétricos, los cuales incorporan elementos de los modelos paramétricos (varianza reducida) y no paramétricos (flexibilidad) [19] [20]. Cuando en este trabajo se apliquen métodos semiparamétricos, el espacio de entrada al modelo se dividirá en dos subespacios diferentes: uno de variables métricas (\mathbf{x}^M) que representan cambios en el nivel de precios; y otro de variables dicotómicas (\mathbf{x}^D) que representan las características propias de la promoción, como por ejemplo la aplicación de un determinado tipo de descuento. Así, el patrón de entrada \mathbf{x} se puede descomponer como $\mathbf{x} = [\mathbf{x}^M \ \mathbf{x}^D]$ y el modelo semiparamétrico obtenido tendrá la siguiente estructura:

$$f(\mathbf{x}) = f_m(\mathbf{x}^M) + f_d(\mathbf{x}^D) \quad (3.3)$$

donde $f_m(\mathbf{x}^M)$ se corresponde con un modelo no paramétrico, mientras que $f_d(\mathbf{x}^D)$ se corresponde con un modelo paramétrico. De esta forma se evita aplicar el procedimiento no paramétrico a todas las características de entrada, consiguiendo que la velocidad de convergencia aumente.

Tras presentar las tres familias de métodos de estimación, se procede al análisis de los métodos de regresión no lineal basados en aprendizaje estadístico y usados en el presente proyecto; a saber, métodos no paramétricos (k-NN, GRNN, MLP y SVM) y semiparamétricos (SVM semiparamétrica).

3.2. Métodos de regresión basados en aprendizaje estadístico

3.2.1. k-NN

El estimador de los k vecinos más próximos (k-NN, del inglés *k-Nearest Neighbors*) fue desarrollado por Fix y Hodges en 1951 [6]. Es un método no paramétrico [39] que estima el valor y de un patrón a partir de las características de los k ejemplos más próximos, utilizando para ello una medida de similitud o distancia [40]. Como medida de distancia se suele usar la norma uno o la distancia euclídea, pero depende en todo caso de la naturaleza de los datos [6]. El valor de k es el parámetro de diseño de este estimador.

Para considerar la influencia de cada ejemplo en el valor a estimar, es lógico asumir que los patrones más cercanos deben tener más influencia, por lo que cada salida y_i se pondera por un valor w_i inversamente proporcional a la distancia entre \mathbf{x} y \mathbf{x}_i . Por tanto, la expresión analítica del estimador k -NN queda como:

$$f(\mathbf{x}) = \frac{\sum_{i=1}^k w_i y_i}{\sum_{i=1}^k w_i} \quad (3.4)$$

donde w_i es una función de ponderación dependiente de la distancia.

3.2.2. Redes Neuronales Artificiales. GRNN y MLP

Las redes neuronales artificiales (ANN, del inglés *Artificial Neural Network*) se describen como modelos matemáticos multiparamétricos no lineales, capaces de aprender a partir de ejemplos y descubrir relaciones que pueden permanecer ocultas en el volumen de datos [41].

Los elementos básicos de la ANN son las neuronas, que son sencillas unidades de procesamiento. El potencial de las ANN proviene de la alta conectividad entre neuronas, normalmente agrupadas en capas: capa de entrada, una o varias capas ocultas y capa de salida. La conexión entre cada par de neuronas tiene asociado un peso (denotado w), que corresponde a un valor numérico obtenido durante una fase de diseño o aprendizaje.

En este proyecto se analizan dos modelos de ANN: GRNN y MLP.

GRNN

La red neuronal de regresión generalizada (GRNN, del inglés *General Regression Neural Network*) es un estimador no paramétrico desarrollado en 1991 por Specht [42], y obtenido a

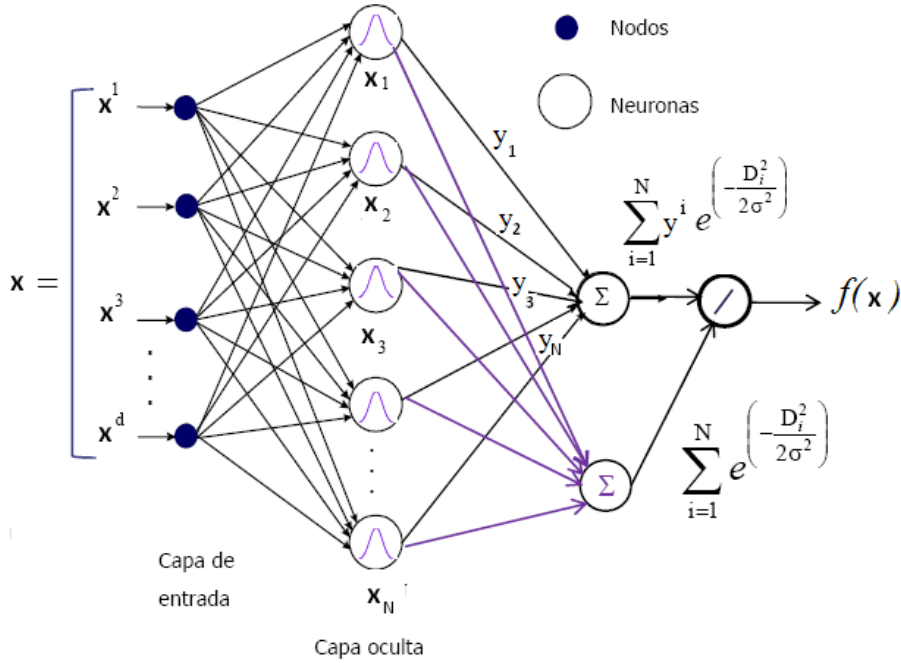


Figura 3.1: *Arquitectura de la GRNN.*

partir de la minimización del error cuadrático sobre el conjunto de ejemplos disponibles. La función $f(\mathbf{x})$ que minimiza este error viene dada por la esperanza de y condicionada a \mathbf{x} [38]:

$$f(\mathbf{x}) = E[y|\mathbf{x}] = \frac{\int yp(\mathbf{x}, y)dy}{\int p(\mathbf{x}, y)dy} \quad (3.5)$$

donde $p(\mathbf{x}, y)$ es la función densidad de probabilidad conjunta de \mathbf{x} e y . Como p no se conoce a priori, se puede estimar utilizando técnicas no paramétricas, como por ejemplo las Ventanas de Parzen (que proporciona estimadores consistentes) con funciones núcleo gaussianas. Sustituyendo el estimador obtenido con las Ventanas de Parzen y la Ecuación (3.2) en la Ecuación (3.5), se obtiene la siguiente expresión para el estimador GRNN:

$$f(\mathbf{x}) = \frac{\sum_{i=1}^N e^{-D_i^2/2\sigma^2} y_i}{\sum_{i=1}^N e^{-D_i^2/2\sigma^2}} = \sum_{i=1}^N h_i y_i \quad (3.6)$$

donde σ es el parámetro de dispersion; $D_i^2 = (\mathbf{x} - \mathbf{x}_i)(\mathbf{x} - \mathbf{x}_i)^T$ es el cuadrado de la distancia Euclídea entre el patrón de entrada \mathbf{x} y el ejemplo de diseño \mathbf{x}_i .

El análisis de la Ecuación (3.6) indica que el valor estimado es una media ponderada de los N valores conocidos y_i , siendo la ponderación no lineal (exponencial dependiente de la distancia Euclídea al punto a estimar). La arquitectura de la GRNN implementa la Ecuación (3.6), como se ilustra en la Figura 3.1. Las N neuronas de la capa oculta son gaussianas esféricas, la información que procesan se comunica a las neuronas de la capa de salida y la red proporciona una salida normalizada.

El parámetro σ es el parámetro libre de la GRNN. Para altos valores de σ , el valor de la salida depende de ejemplos alejados, y la salida es una estima suavizada del valor real. Por el contrario, si σ se hace muy pequeño, la red tiende a estimar el valor asociado al ejemplo más próximo a \mathbf{x} [42].

MLP

El perceptrón multicapa o MLP (del inglés *Multi Layer Perceptron*) es una red neuronal cuyas neuronas están agrupadas en capas: una capa de entrada, una o varias capas ocultas y una capa de salida, tal y como se ilustra en la Figura 3.2.

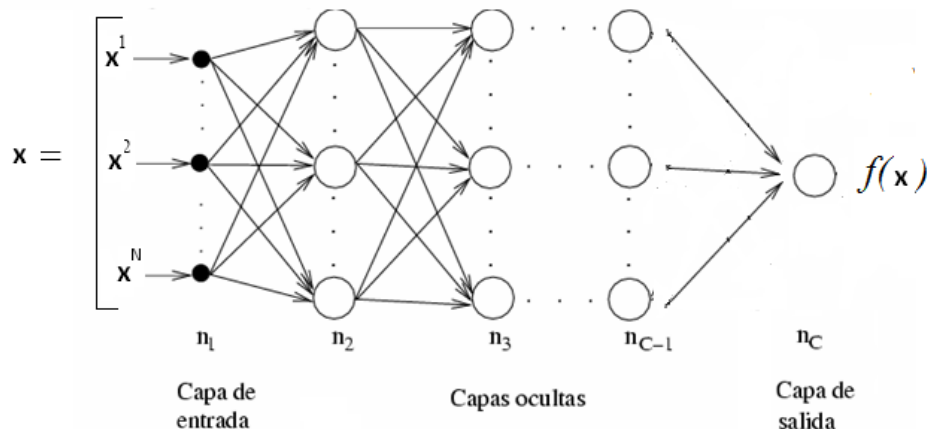


Figura 3.2: Arquitectura del MLP con $n_c - 2$ capas ocultas.

El número de neuronas de la capa de entrada y salida están determinadas por la naturaleza de los datos: existen tantas neuronas de entrada como dimensiones tenga el patrón de entrada \mathbf{x} , y tantas neuronas de salida como variables a estimar (dimensión de la variable y). Así pues, el diseñador del modelo de red debe elegir el número de capas ocultas y de neuronas en cada capa oculta. Las neuronas de la capa oculta implementan funciones globales (denominadas funciones de activación), que pueden ser de tipo lineal o sigmoideal (con salida acotada, por ejemplo en el intervalo $[0, 1]$). Si la función de activación es de tipo sigmoideal, el MLP se considera un

aproximador universal, es decir, con una única capa oculta se puede aproximar cualquier función continua y acotada [43].

Durante la fase de aprendizaje se determinan los pesos sinápticos $\{w\}$ asociados a las conexiones entre neuronas. Existen varios algoritmos de aprendizaje, todos ellos basados en optimizar una función de coste. El algoritmo más sencillo es el algoritmo de retropropagación (en inglés *back-propagation*) [38]. Se trata de un algoritmo iterativo que parte de una solución inicial y optimiza una función de coste (error cuadrático) aplicando una técnica de descenso por gradiente (utilizando solo primeras derivadas) para cada uno de los pesos sinápticos. El algoritmo compara la salida de la red con la salida deseada y la señal de error se propaga desde la capa de salida a la de entrada, permitiendo ajustar todos los parámetros del MLP a través de un proceso iterativo. Si la función a optimizar no es convexa en el espacio de búsqueda, es posible que los pesos sinápticos proporcionados por el algoritmo correspondan a un mínimo local (no óptimo) de la función a optimizar. Para intentar resolver este problema se puede incorporar información sobre las segundas derivadas, tal y como hace el algoritmo de Levenberg-Marquardt usado en este proyecto [44].

3.2.3. SVM

Las máquinas de vectores soporte (SVM, del inglés *Support Vector Machine*) son métodos de aprendizaje estadístico denominados métodos núcleo. Los últimos años han supuesto el auge de las SVM debido a las ventajas que presentan: escalan relativamente bien con patrones de entrada de alta dimensión, son robustas ante muestras atípicas y la función a optimizar no tiene mínimos locales en el espacio transformado. La idea fundamental es realizar una transformación no lineal $\phi(\cdot)$ a otro espacio de características, generalmente de mayor dimensión que el espacio de representación original, y en el cual se puede resolver el problema utilizando técnicas de regresión lineal.

Si la transformación $\phi(\cdot)$ es no lineal, la función $f(\cdot)$ será no lineal en el espacio original. A continuación se presenta la formulación utilizando una función $\phi(\cdot)$ lineal y posteriormente su extensión al caso no lineal.

Se desea aproximar la función de regresión $f(\cdot)$ por un hiperplano regresor (definido por el vector \mathbf{w}) en el espacio de entrada:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (3.7)$$

Para conseguir una solución robusta, la regresión SVM optimiza la función de coste ϵ -insensible de Vapnik (L_ϵ) [45], mostrada en la Figura 3.3(a), que incorpora una región de insensibilidad (de anchura ϵ) a los errores en la proximidad del hiperplano de regresión. De forma matemática se expresa como:

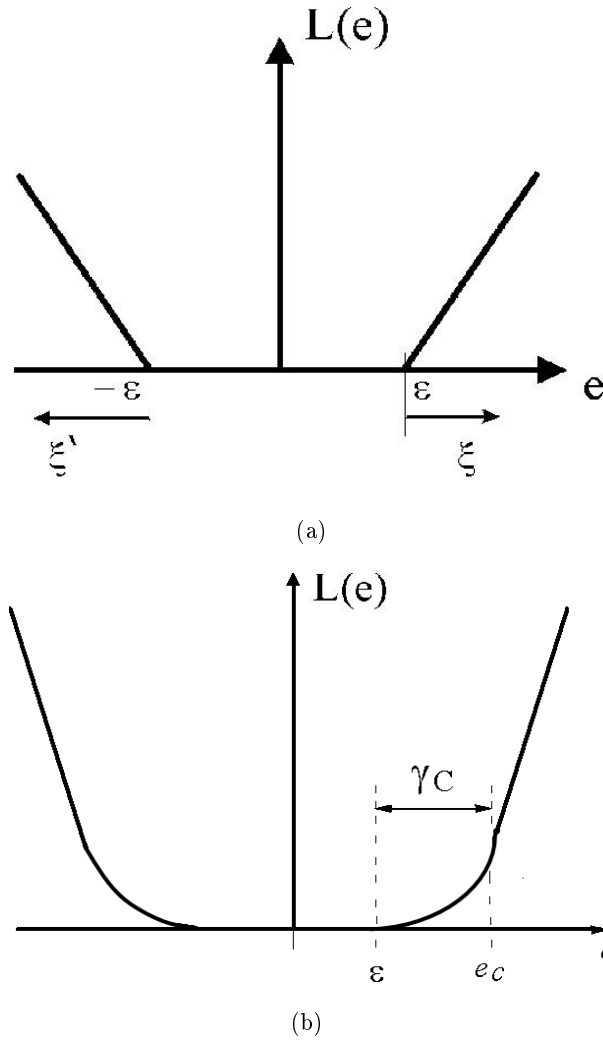


Figura 3.3: *Función de coste: (a) ϵ -insensible de Vapnik; (b) ϵ -Huber.*

$$L_{\epsilon}(y_i, f(\mathbf{x}_i)) = \begin{cases} |y_i - f(\mathbf{x}_i)| - \epsilon & \text{si } |y_i - f(\mathbf{x}_i)| \geq \epsilon \\ 0 & \text{en otro caso} \end{cases} \quad (3.8)$$

donde ϵ es el parámetro de insensibilidad, de modo que se ignoran errores menores que $|\epsilon|$. El coste ϵ -insensible de Vapnik presenta algunas limitaciones debido a que es una función lineal. Se ha propuesto utilizar otra función de coste, denominada coste ϵ -Huber (véase Figura 3.3(b)) [46], que combina tres regiones: una región insensible, una cuadrática y una lineal. La función ϵ -Huber tiene un parámetro más que la función ϵ -insensible de Vapnik, el parámetro γ , que controla el límite entre la zona cuadrática y la zona lineal.

En la Figura 3.3(a) se representa la Ecuación (3.8) y en la Figura 3.4 se muestra el hiperplano regresor, la zona de insensibilidad y las variables ξ y ξ' , denominadas *slack variables*, que definen

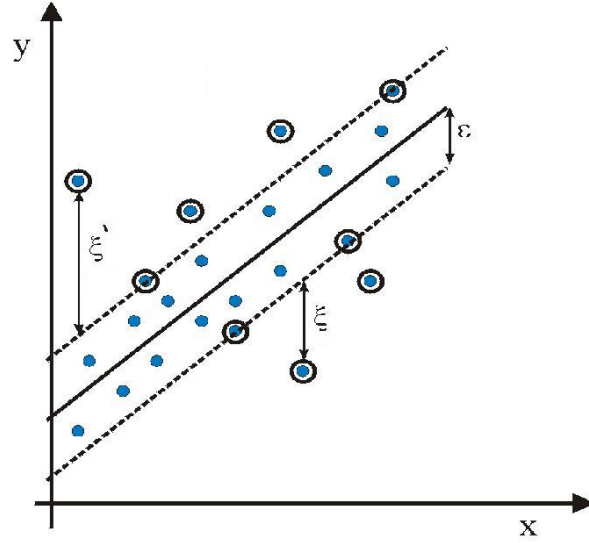


Figura 3.4: *Hiperplano regresor en la SVM y concepto de ϵ -insensibilidad.*

el error cometido según la función L_ϵ . Los ejemplos que se encuentran en la frontera del margen de insensibilidad o exteriores a él (señalados con un círculo negro en la Figura 3.4), se conocen como vectores soporte.

El parámetro de insensibilidad ϵ no tiene soporte compacto (puede tomar cualquier valor positivo), por lo que su selección conlleva un extenso barrido si se desconoce a priori la exactitud de la aproximación. Como alternativa se ha propuesto utilizar el método ν -SVM (considerado en este proyecto), que permite calcular automáticamente el valor de ϵ , a través de un nuevo parámetro ν cuyo rango está acotado en el intervalo $(0, 1)$ [47].

El algoritmo ν -SVM diseña el modelo optimizando la siguiente función para el coste ϵ -insensible:

$$L = \frac{1}{2} \|\mathbf{w}\|^2 + C \left(\nu \epsilon + \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi'_i) \right) \quad (3.9)$$

donde C es el parámetro de regularización (escogido a priori) que controla el compromiso entre error y suavidad de la superficie, y ν es el parámetro que establece la proporción entre el número de vectores soporte y el total de ejemplos de diseño. La optimización de la función L está sujeta a las siguientes restricciones:

$$(\mathbf{w}^T \mathbf{x}_i + b) - y_i \leq \epsilon + \xi_i \quad (3.10)$$

$$y_i - (\mathbf{w}^T \mathbf{x}_i + b) \leq \epsilon + \xi'_i \quad (3.11)$$

$$\xi_i, \xi'_i \geq 0 \quad (3.12)$$

$$\epsilon \geq 0 \quad (3.13)$$

con $i = 1, \dots, N$. La solución final a este problema de optimización permite obtener el siguiente regresor lineal:

Para encontrar los parámetros del regresor lineal de la Ecuación (3.7), se construye el Lagrangiano a minimizar, siendo α , α' , η , η' y β los multiplicadores de Lagrange:

$$\begin{aligned} L = & \frac{1}{2} \|\mathbf{w}\|^2 + C\nu\epsilon + \frac{C}{N} \sum_{i=1}^N (\xi_i + \xi'_i) - \beta\epsilon - \sum_{i=1}^N (\eta_i \xi_i + \eta'_i \xi'_i) - \\ & - \sum_{i=1}^N \alpha_i (\xi_i + y_i - \mathbf{w}^T \mathbf{x}_i - b + \epsilon) - \sum_{i=1}^N \alpha'_i (\xi'_i + \mathbf{w}^T \mathbf{x}_i + b - y_i + \epsilon) \end{aligned} \quad (3.14)$$

Para resolver el problema de optimización se debe encontrar el punto de silla de L , lo que implica minimizar la Ecuación (3.14) respecto a las variables primales (\mathbf{w} , ϵ , b , ξ_i y ξ'_i), y maximizar respecto a las variables duales (α_i , α'_i , η_i , η'_i y β). Así, derivando respecto a las variables primales e igualando a cero, se obtienen las siguientes cinco ecuaciones:

$$\mathbf{w} = \sum_{i=1}^N (\alpha'_i - \alpha_i) \mathbf{x}_i \quad (3.15)$$

$$C\nu - \sum_{i=1}^N (\alpha'_i + \alpha_i) - \beta = 0$$

$$\sum_{i=1}^N (\alpha_i - \alpha'_i) = 0$$

$$\frac{C}{N} - \alpha_i - \eta_i = 0$$

$$\frac{C}{N} - \alpha'_i - \eta'_i = 0$$

Sustituyendo la Ecuación (3.15) en la Ecuación (3.7) se obtiene:

$$f(\mathbf{x}) = \left[\sum_{i=1}^N (\alpha'_i - \alpha_i) \mathbf{x}_i^T \right] \mathbf{x} + b \quad (3.16)$$

Las variables α'_i y α_i sólo serán distintas de cero cuando \mathbf{x}_i se encuentre en el borde o en el exterior de la zona de insensibilidad. Obsérvese que el valor a estimar se puede obtener a partir

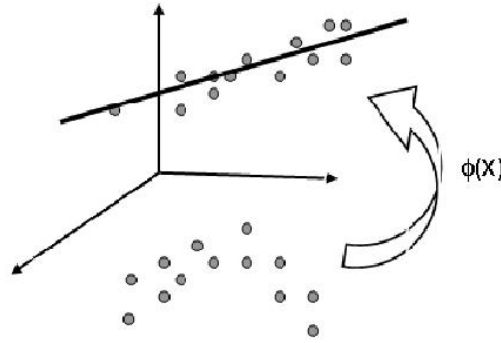


Figura 3.5: Transformación $\phi(\cdot)$ a un espacio de características de mayor dimensión.

del producto escalar entre los datos, sin necesidad de calcular explícitamente \mathbf{w} . Esto permite formular una extensión no lineal usando *funciones núcleo*. La idea es que el modelo lineal puede no ser adecuado en el espacio original, existiendo una función no lineal $\phi(\mathbf{x})$ que transforme los ejemplos a un espacio de características de mayor dimensionalidad en el que un modelo lineal sí sea adecuado (véase la Figura 3.5) [46]. En la formulación SVM, el conocimiento explícito de la función $\phi(\cdot)$ se puede evitar usando un *Núcleo de Mercer* $K(\mathbf{x}, \mathbf{x}_i)$, que se relacionan del siguiente modo con el producto escalar en el espacio transformado:

$$K(\mathbf{x}, \mathbf{x}_i) = \phi(\mathbf{x})^T \phi(\mathbf{x}_i) \quad (3.17)$$

En este proyecto se emplea un núcleo de tipo RBF gaussiano cuando se considera como esquema ν - SVM, de modo que:

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right) \quad (3.18)$$

Sustituyendo la función núcleo y las ecuaciones resultantes de optimizar respecto a las variables primales en la expresión de la Ecuación (3.14), se obtiene el siguiente problema dual:

$$\max_{\alpha'_i, \alpha_i} \sum_{i=1}^N (\alpha_i - \alpha'_i) y_i - \frac{1}{2} \sum_{i,j=1}^N (\alpha'_i - \alpha_i)(\alpha'_j - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j) \quad (3.19)$$

sujeto a

$$\sum_{i=1}^N (\alpha_i - \alpha'_i) = 0 \quad (3.20)$$

$$\sum_{i=1}^N (\alpha_i + \alpha'_i) \leq C\nu$$

$$\alpha_i, \alpha'_i \in \left[0, \frac{C}{N}\right]$$

De modo que el estimador final sigue la siguiente expresión:

$$f(\mathbf{x}) = \sum_{i=1}^N (\alpha'_i - \alpha_i)K(\mathbf{x}, \mathbf{x}_i) + b \quad (3.21)$$

donde b se obtiene a partir de las Ecuaciones (3.10) y (3.11) en los ejemplos en los que se cumpla que $\xi_i = \xi'_i = 0$.

El esquema ν -SVM merece una consideración especial en lo que se refiere a la búsqueda de los parámetros libres. En la Sección 2 se han explicado los diferentes esquemas SVM con los que se trabaja en este proyecto: SVM con núcleo RBF, ν -SVM y SVM semiparamétrica. Esto implica que, en función del esquema SVM usado para estimar el volumen de unidades vendidas, el número de parámetros libres será distinto.

Si el problema a resolver es lineal, los parámetros libres son aquéllos asociados a la función de coste:

- ϵ es el límite de error que no se penaliza.
- C es el parámetro de regularización o compromiso entre el error empírico y la suavidad de la solución.
- γ es la constante que fija el umbral de paso de la función de coste cuadrática a la zona lineal cuando se utiliza el coste ϵ -Huber.
- ν es un límite superior de la fracción de errores de margen y un límite inferior de la fracción de vectores de soporte. De esta definición se hace evidente que el papel de σ y ϵ están relacionados.

Si el problema a resolver es no lineal, adicionalmente se tienen los parámetros asociados al núcleo de Mercer (parámetro de suavizado σ en el núcleo gaussiano).

Cuando se estima el número de unidades vendidas considerando el esquema ν -SVM semiparamétrico, otro parámetro libre relacionado con el núcleo es el parámetro de compensación to (del inglés *trade-off*). Este parámetro permite obtener el núcleo semiparamétrico como suma de dos núcleos: uno lineal, ponderado por to , y otro de tipo RBF.

3.3. Figuras de mérito y técnicas de evaluación

Los métodos de aprendizaje permiten construir modelos a partir de un conjunto de datos. Una primera aproximación para evaluar la calidad del modelo es usar el conjunto de datos usado para entrenar el modelo, y proporcionar una figura de mérito. Sin embargo, este procedimiento no garantiza la generalización, porque premia los ejemplos que más se ajustan al conjunto de diseño. Otra aproximación más adecuada es evaluar el modelo sobre otro conjunto no utilizado durante el diseño.

3.3.1. Figuras de mérito

Las figuras de mérito permiten comparar las prestaciones de distintos modelos en base a la diferencia entre las salidas deseadas y las salidas estimadas. En este proyecto se han considerado diferentes figuras de mérito:

Absolutas. En general, a no ser que se indique lo contrario, las prestaciones son tanto mejores cuanto más reducido es el valor de la figura de mérito.

- MAE (del inglés *Mean Absolute Error*):

$$MAE = \frac{1}{N} \sum_{i=1}^N |f(\mathbf{x}_i) - y_i| \quad (3.22)$$

- MSE (del inglés *Mean Square Error*):

$$MSE = \frac{1}{N} \sum_{i=1}^N (f(\mathbf{x}_i) - y_i)^2 \quad (3.23)$$

Como medida de dispersión se presenta:

- Anchura del *IC* (Intervalo de confianza):

Para una variable aleatoria, corresponde al rango de valores entre los que ésta se encuentra con una determinada probabilidad o nivel de confianza. El *IC* se suele expresar como $[IC_{inf}, IC_{sup}]$, donde IC_{inf} es el límite inferior y IC_{sup} es el límite superior. La anchura del *IC* indica la dispersión de la variable aleatoria: cuanto menor sea esta anchura, menor es la dispersión obtenida.

Relativas. Sus valores varían entre 0 y 1.

- RMAE (del inglés *Root Mean Absolute Error*):

$$RMSE = \frac{1}{N} \sum_{i=1}^N \left| \frac{f(\mathbf{x}_i) - y_i}{y_i} \right| \quad (3.24)$$

- RRMSE (del inglés *Root Mean Square Error*):

$$RRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \frac{(f(\mathbf{x}_i) - y_i)^2}{y_i}} \quad (3.25)$$

- *Agreement Index* o D de Willmott [48], un valor de $D = 1$ indica que el valor estimado es el mejor posible:

$$Willmott's D = 1 - \frac{\sum_{i=1}^N (f(\mathbf{x}_i) - y_i)^2}{\sum_{i=1}^N (|f(\mathbf{x}_i) - \bar{y}| + |y_i - \bar{y}|)^2} \quad (3.26)$$

\bar{y} es el valor medio de los ejemplos.

3.3.2. Validación cruzada

Los métodos de validación cruzada son métodos estadísticos para estimar el error de generalización. Fueron formulados inicialmente por Seymour Geisser [49] y permiten evaluar las prestaciones de un modelo y comparar varios modelos para identificar el mejor. En este proyecto se aplica una técnica de validación cruzada para encontrar los parámetros libres de cada método (k-NN, SVM, GRNN y MLP). La técnica de validación cruzada consiste en separar un conjunto de ejemplos en dos subconjuntos: uno para diseñar el modelo (conjunto de entrenamiento) y otro para comprobar su comportamiento y prestaciones de generalización (conjunto de validación) [47]. Existen tres técnicas diferentes de validación cruzada:

- **Holdout.** Es el caso más sencillo en el que se reserva un porcentaje aleatorio de ejemplos como conjunto de validación y el resto de ejemplos se usa como conjunto de entrenamiento. En general, el porcentaje de validación suele ser del 30%. Es habitual, no obstante, encontrar porcentajes diferentes para realizar la partición [50].

- **K-fold.** En este caso, el conjunto de ejemplos se divide aleatoriamente en K subconjuntos del mismo tamaño: un subconjunto se reserva para validación y los ejemplos de los $K-1$ subconjuntos restantes se usan para diseñar el modelo. El proceso se repite K veces, dejando cada vez uno de los K subconjuntos diferentes como validación. La generalización del modelo se estima promediando las prestaciones de los K modelos.
- **Leave-one-out (LOO).** Es un caso particular de la técnica K -fold, donde K coincide con el número de ejemplos disponibles.

En este proyecto se ha utilizado la técnica LOO para estimar los parámetros libres de los modelos citados anteriormente (k-NN, GRNN, MLP y SVM), por sus ventajas, cuando el número de ejemplos disponible es escaso. No obstante, la carga computacional de la técnica LOO es superior a la de otras técnicas de validación, ya que la validación se realiza tantas veces como número de ejemplos hay.

3.3.3. Evaluación con remuestreo *bootstrap*

Con el objetivo de comprobar si existen diferencias estadísticamente significativas entre las prestaciones de los distintos métodos de aprendizaje máquina utilizados para modelar las respuestas promocionales en un conjunto de datos bajo estudio, se recurre a la utilización de una técnica no paramétrica denominada remuestreo *bootstrap*, especialmente indicada en casos con un número de observaciones reducido y de distribución desconocida (como es el caso en este proyecto). Asumiendo que se dispone de N observaciones, esta técnica permite construir una aproximación a la distribución de las medidas de prestaciones realizando un muestreo aleatorio con reposición de las N observaciones disponibles como medidas de la figura de mérito. Puesto que el procedimiento de remuestreo es con reposición, es posible que en el conjunto de observaciones remuestreadas haya observaciones repetidas y, por tanto, dicho conjunto también puede en general no contener varios de los ejemplos. El conjunto de N observaciones obtenidas a partir del conjunto original mediante muestreo con reposición se conoce como *remuestra*, y se repite un número prefijado de B veces.

Debido a las características de las bases de datos tratadas en el presente proyecto, se ha decidido utilizar este método no paramétrico para contrastar las unidades vendidas con distintos modelos. El remuestreo *bootstrap* fue propuesto inicialmente como una aproximación razonable de la distribución estadística de una variable aleatoria ya que se ha observado previamente que la distribución de los residuos de estimación presenta en general colas largas y distribuciones que no siempre se asemejan a la gaussiana, y ha sido aplicado con ventajas demostradas en una amplia variedad de situaciones [51].

En este trabajo se presenta el remuestreo *bootstrap* como herramienta para describir en detalle

la distribución estadística de las prestaciones de los modelos promocionales basados en aprendizaje, tanto de forma individual, para caracterizar la salida de cada modelo, como de forma pareada, para contrastar las prestaciones de los diferentes modelos analizados. Este test se aplica sobre los errores obtenidos, \mathbf{e}_i , al estimar el número de unidades vendidas de cada modelo, o bien sobre algunos parámetros que caracterizan el mérito y cuya distribución no es sencilla de calcular en teoría de forma analítica (por ejemplo, el límite superior de un intervalo de confianza)..

En una primera aproximación, se propone el remuestreo *bootstrap* para contrastar las salidas obtenidas de forma individual en cada uno de los métodos. El procedimiento seguido se describen a continuación:

- A partir de la muestra original de errores de predicción para un modelo dado, $\mathbf{e} = [e_1, e_2, \dots, e_N]$, se seleccionan B remuestras *bootstrap* independientes, $\mathbf{e}_1^*, \mathbf{e}_2^*, \dots, \mathbf{e}_B^*$, cada una compuesta por N datos tomados de la muestra original \mathbf{e} con reposición.
- Para cada remuestra obtenida, se calcula el valor de un determinado estadístico θ que se utiliza como estimador del parámetro poblacional, en cuyo estudio se está interesado, utilizando un operador de cálculo t , esto es:

$$\hat{\theta}_b^* = t(\mathbf{e}_b^*), \quad b = 1, 2, \dots, B \quad (3.27)$$

En este trabajo, y como se verá en el Experimento 1, $\hat{\theta}$ es una medida de error (MAE) o de dispersión, como se describirá más adelante.

- Se construye una distribución muestral de las réplicas del estadístico $\hat{\theta}^*$, que representa una aproximación a la verdadera distribución de probabilidad para el estadístico $\hat{\theta}$.

Se determina de este modo la distribución muestral de un estadístico sin haber realizado suposiciones sobre la distribución teórica a la que ésta se ajusta y sin manejar fórmulas analíticas para determinar los correspondientes parámetros de esa distribución. Sobre dicha distribución muestral se puede calcular su intervalo de confianza con una probabilidad del 95%. Si el valor obtenido de forma empírica (MAE obtenido a partir de las salidas deseadas y de la estimadas) pertenece a dicho *IC*, se puede asumir que el estimador obtenido de forma empírica no presenta sesgo debido a un posible sobreajuste.

Para poder comparar las prestaciones de dos métodos diferentes o la inclusión de las variables dicotómicas es necesario realizar un remuestreo *bootstrap* pareado. La principal diferencia que existe con el remuestreo *bootstrap* individual es que, en este caso, se toman las mismas remuestras de dos muestras procedentes de diferentes métodos de estimación basados en aprendizaje estadístico. Los pasos seguidos se resumen a continuación:

- Se parte de las medidas de error producidas por dos modelos de aprendizaje diferentes, \mathbf{r} y \mathbf{s} , dadas por:

$$\mathbf{r} = [r_1, r_2, \dots, r_N]$$

$$\mathbf{s} = [s_1, s_2, \dots, s_N]$$

y se construyen las remuestras para los incrementos de dichas medidas de error:

$$\Delta = \mathbf{r} - \mathbf{s} = [r_1 - s_1, r_2 - s_2, \dots, r_N - s_N]$$

siendo por tanto las remuestras poblacionales las siguientes:

$$\Delta_1^*, \Delta_2^*, \dots, \Delta_B^*$$

- Para dichas remuestras, se calcula el valor de un determinado estadístico θ que se utiliza como estimador del parámetro poblacional, en cuyo estudio se está interesado, esto es:

$$\hat{\theta}_b^* = t(\Delta_b^*), \quad b = 1, 2, \dots, B \quad (3.28)$$

- Se construye una distribución empírica del estadístico que representa una aproximación de la verdadera distribución de probabilidad para el estadístico pareado.

En el presente trabajo, cuando se realiza remuestreo paramétrico pareado entre dos modelos (*modelo₁* y *modelo₂*) se comparan los resultados obtenidos con tres estadísticos diferentes:

$$\Delta MAE = MAE(modelo_1) - MAE(modelo_2) \quad (3.29)$$

$$\Delta IC = \Delta IC(modelo_1) - \Delta IC(modelo_2) \quad (3.30)$$

$$\Delta IC_{Sup} = IC_{sup}(modelo_1) - IC_{sup}(modelo_2) \quad (3.31)$$

donde IC hace referencia a los valores del intervalo de confianza obtenidos con un nivel de confianza del 95 %; y $\Delta IC(modelo_i) = IC_{sup}(modelo_i) - IC_{inf}(modelo_i)$.

Si se tiene en cuenta las características de los datos usados en la parte experimental, el remuestreo *bootstrap* aporta dos ventajas:

- Como procedimiento no paramétrico, no hace asunción alguna acerca de la distribución estadística de las muestras observadas.
- Las bases de datos usadas para el diagnóstico del modelo presentan baja densidad muestral. El remuestreo *bootstrap* es un procedimiento robusto cuando se dispone de pocas muestras, dentro de un límite suficientemente representativo de la poblacional original.

Bases de datos y modelos de respuesta promocional

Este capítulo presenta los materiales utilizados en el presente proyecto. En particular, las bases de datos utilizadas (promociones en leche y cerveza) y los modelos de respuesta promocional al que se ajustan los diferentes métodos de aprendizaje estadístico.

En la Sección 4.1 se describen las dos bases de datos usadas, leche y cerveza. Los modelos de respuesta promocional se describe en la Sección 4.2. En la Sección 4.3 se realiza un análisis descriptivo de ambas bases de datos a nivel de producto.

4.1. Bases de datos

En este apartado se describe la estructura de dos bases de datos reales de las categorías de leche y cerveza. Se dispone de las ventas realizadas en un único establecimiento minorista (supermercado) a lo largo de un año y excluyendo los fines de semana. El número de registros en ambas bases de datos es de 304 ejemplos, correspondientes a los días en los que se dispone de transacciones de ventas realizadas en el supermercado, agregadas en 43 semanas. A lo largo de este trabajo se estima el número de unidades vendidas semanales, y no diarias, para no tener excesiva influencia de la fuerte estacionalidad semanal presente en los datos.

4.1.1. En la categoría de leche

La primera base de datos analizada es la correspondiente a la categoría de leche. Debido a que es un producto de consumo diario, se ha decidido analizar en este proyecto su reacción en términos de unidades vendidas frente a las promociones realizadas.

Se ha analizado el comportamiento de 6 productos de leche, por lo que van a existir 6 modelos

MODELO	PRODUCTO
Modelo 1	Asturiana
Modelo 2	Ato
Modelo 3	MMDD
Modelo 4	Pascual Calcio
Modelo 5	Pascual Clásica
Modelo 6	Puleva Calcio

Tabla 4.1: *Productos analizados en la categoría de leche y su correspondencia en nomenclatura con los modelos de datos.*

diferentes, correspondientes a cada uno de los productos, tal y como se indica en la Tabla 4.1.

4.1.2. En la categoría de cerveza

Tras revisar la literatura existente, se ha decidido analizar también el comportamiento que presenta la categoría cerveza en las promociones de ventas debido a su fuerte carácter estacional [21]. Existen estudios que demuestran la importancia de diferentes tipos de promociones en la categoría de cerveza, como se indica en las Tablas 4.2 y 4.3. Estas tablas muestran la distribución ponderada. Se entiende por distribución ponderada el porcentaje de las tiendas en las que se encuentra un producto determinado sobre el total de tiendas. Indica la penetración real de un producto, puesto que también tiene en cuenta la importancia y calidad de cada establecimiento [21]. Si no se trabajase con una distribución ponderada, podría ocurrir que un tipo de promoción tuviera un alto porcentaje de distribución, pero en tiendas pequeñas o poco importantes para las ventas de la clase de producto analizado.

Al igual que para los datos de la categoría leche, en el presente trabajo se dispone de una base de datos previamente preprocesada sobre la que analizar el comportamiento promocional de 14 productos diferentes. En la Tabla 4.4 se muestra qué producto se corresponde a cada modelo de forma que, a lo largo del estudio, se hablará en términos de modelos y no de productos. Para algunas marcas se han analizado diferentes formatos; por ejemplo, para Amstel se estima el número de unidades vendidas para la lata de 33 centilitros y para el *pack* de 6 botellines de 25 centilitros. El número de observaciones coincide con el registrado en leche, 304 ejemplos correspondientes a los días en los que el supermercado tiene ventas, y agregados en 43 semanas.

4.2. Modelos de respuesta promocional

Los modelos de respuesta promocional planteados en este proyecto pretenden estudiar la variación de las ventas en función del índice de precios y de variables relativas al tipo de promoción.

Productos	Reducción temporal de precio	Folleto exposición espacial	Exposición espacial	Folleto	Promoción regalo directo	Resto de promos
1º Cerveza	99	44	79	67	3	92
2º Bebidas refrescantes	98	58	91	71	15	97
3º Paté	98	18	61	45	52	99
4º Café tostado	96	26	69	57	19	96
5º Alimentos para perros	92	23	47	45	11	92
6º Snacks	92	33	77	46	22	67
7º Sopas	91	10	24	33	7	88
8º Galletas maría	90	23	45	45	26	86
9º Chocolate	90	16	53	40	26	85
10º Aceite	88	28	60	56	1	82
11º Alimentos para gatos	86	8	29	28	2	88
12º Quesos lonchas	85	15	31	41	4	92
13º Whisky	85	24	55	60	50	86
14º Queso porciones	83	15	40	38	16	75
15º Licores	77	9	20	45	30	73

Tabla 4.2: *Distribuciones ponderadas de los diversos tipos de promociones en diferentes categorías de productos (tomado de [21]).*

	Folleto + Exposición especial	Exposición especial	Folleto	Resto de promos	Reducción temporal de precios	Regalo directo
TOTAL ESPAÑA	44	79	67	97	99	3
100-399 M2	19	32	42	63	81	0
400-999 M2	31	55	42	73	86	0
1000 - 2499 M2	27	66	35	69	73	0
HIPERMERCADO	44	79	67	97	99	3

Tabla 4.3: *Distribuciones ponderadas de los tipos de promociones de cerveza (tomado de [21]).*

Para el desarrollo del modelo se distinguen dos tipos de variables:

- **Variable endógena o explicada.** Es la variable a modelar y corresponde con el volumen de ventas de un determinado producto en un establecimiento y durante una semana determinada.
- **Variables exógenas o explicativas.** Son las variables que explican el comportamiento de las unidades vendidas. En este proyecto se consideran tres tipos de variables, las que indican la reducción del precio, la que indica si ha habido descuento directo y la variable

MODELO	PRODUCTO
Modelo 1	Amstel 25 por 6
Modelo 2	Amstel 33
Modelo 3	Bavaria 33
Modelo 4	Cruzcampo 33
Modelo 5	Estrella 25 por 6
Modelo 6	Estrella 25 por 12
Modelo 7	Estrella 33
Modelo 8	Heineken 25 por 6
Modelo 9	MMDD 33
Modelo 10	San Miguel 25 por 6
Modelo 11	San Miguel 25 por 12
Modelo 12	Voll Damm 25 por 6
Modelo 13	Xibeca 25 por 6
Modelo 14	Xibeca 33

Tabla 4.4: *Marcas analizadas en la categoría de cerveza y su correspondencia en términos de modelos.*

que indica la estacionalidad. Puesto que se pretende estudiar cómo varían las ventas en función del descuento en precio realizado, existen dos tipos de variables explicativas: el precio promocional como tal, o bien la variación existente entre el precio promocional y otro precio. En ocasiones, este otro precio se conoce como *precio de referencia*, y puede utilizarse cualquier precio que el consumidor considera como base para la comparación. Se puede tratar por tanto de un precio subjetivo, y puede ser dinámico en función del entorno [4].

De forma común, el descuento relativo se calcula como:

$$D_r(i, t) = \frac{P_{ref}(i, t) - P_{prom}(i, t)}{P_{ref}(i, t)} \quad (4.1)$$

donde $D_r(i, t)$ es el descuento relativo del producto i en la semana t ; $P_{ref}(i, t)$ es el precio de referencia del producto i en la semana t ; $P_{prom}(i, t)$ es el precio promocional del producto i en la semana t ; y t está expresado en semanas. En lugar del descuento relativo, en este proyecto se trabaja con el índice de precios, dado por:

$$I_{pr}(i, t) = \frac{P_{prom}(i, t)}{P_{reg}(i, t)} \quad (4.2)$$

donde $I_{pr}(i, t)$ es el índice de precio del producto i en la semana t ; y $P_{reg}(i, t)$ es el precio regular del producto i en la semana t .

El índice de precios expresa en tanto por uno la variación existente entre el precio promocional y el precio regular, y toma el valor 1 si ambos precios son iguales. Este índice permite una comparación más clara de la magnitud de los descuentos realizados a lo largo del año, y por ello estos índices frecuentemente se consideran como variables de entrada más informativas que el precio promocional.

Con objeto de incluir el impacto de los tipos de descuentos promocionales de todos y cada uno de los productos competidores sobre las ventas de un determinado producto, además del índice de precios se han considerado otras variables exógenas que se describen a continuación. Se ha utilizado la variable descuento directo (reducción en precio), variable dicotómica que toma el valor 1 si existe descuento directo y 0 en caso contrario. Otra variable exógena es la línea de base, la cual permite identificar posibles estacionalidades en las unidades vendidas a lo largo del año objeto de estudio. Para simplificar el estudio de la misma, y tratarla como una variable dicotómica más, sólo se trabaja con dos posibles tramos, identificando cada uno de ellos con 0 ó 1, lo que se corresponde de forma natural con la posibilidad de distinguir la temporada baja de la temporada alta.

En la Figura 4.1(a) se muestran los índices de precios para cada uno de los modelos de la categoría leche, y en la Figura 4.1(b) las unidades vendidas en cada semana. En la Figura 4.2(a) se muestran los índices de precios para cada uno de los modelos de la categoría de cerveza, y en la Figura 4.2(b) las unidades semanales vendidas para cada uno de los 14 productos de cerveza analizados.

Existen ciertas características comunes del modelo de respuesta promocional para las categorías de leche y cerveza: la entrada se corresponde con diferentes combinaciones de índices de precios, y a la salida del modelo se obtiene una estimación del número de unidades vendidas para cada producto. Se pretende que la salida obtenida sea lo más parecida a la salida deseada sin que exista sobreajuste. Un modelo de respuesta promocional se puede expresar como:

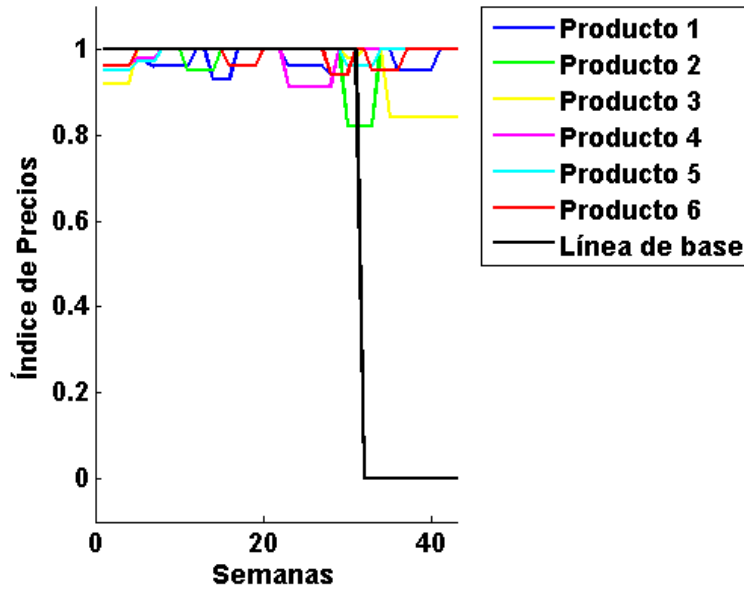
$$y(i, t) = f(\mathbf{x}^M(i, t), \mathbf{x}^D(i, t), LB(t)) \quad (4.3)$$

siendo $y(i, t)$ el número de unidades vendidas del producto i en la semana t ;

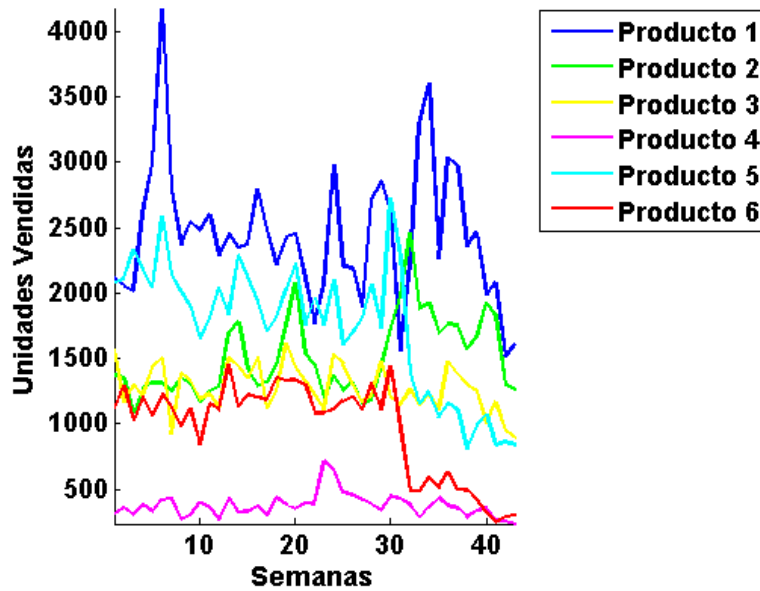
$\mathbf{x}^M(i, t) = [IP^1(i, t), \dots, IP^{n_m}(i, t)]$ es un vector con los índices de precios del producto i en la semana t , con $n_m = 6$ para leche y $n_m = 14$ para cerveza; $\mathbf{x}^D(i, t)$ es la de variable dicotómica descuento directo asociada al producto i en la semana t ; y $LB(t)$ es la línea de base en la semana t .

4.3. Análisis a nivel de producto

En primer lugar se analiza cada uno de los productos en términos de unidades vendidas e índice de precios, con el objetivo de comprobar su comportamiento ante las promociones. En

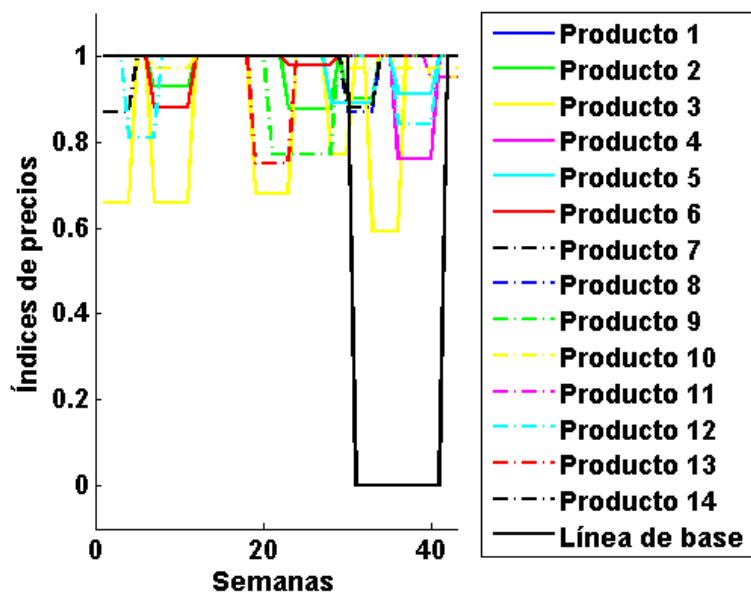


(a)

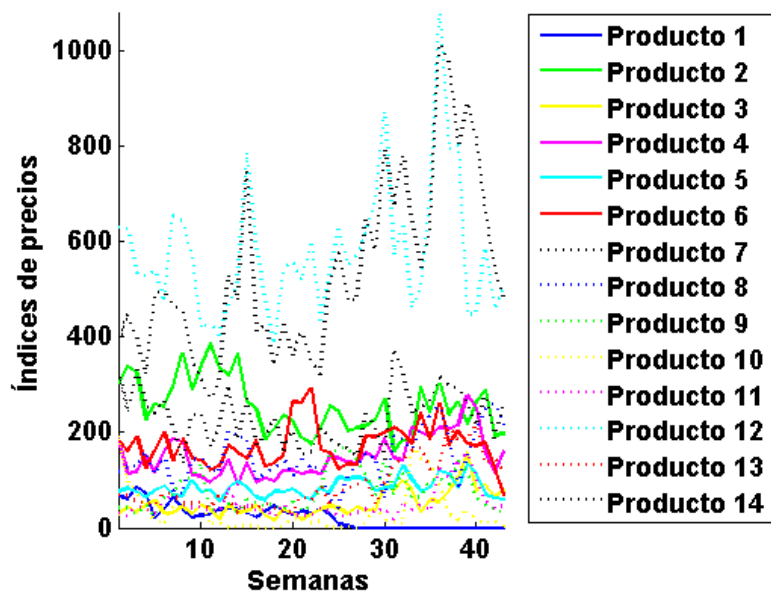


(b)

Figura 4.1: Evolución temporal de los índices de precios y las unidades vendidas en la categoría de leche (6 productos): (a) índice de precios; (b) unidades vendidas.



(a)



(b)

Figura 4.2: Evolución temporal de los índices de precios y las unidades vendidas en la categoría de cerveza (14 modelos): (a) índice de precios; (b) unidades vendidas.

segundo lugar, y debido a la existencia de productos de una misma marca, se realiza un análisis comparativo de los mismos. La marca del distribuidor se estudia en último lugar.

Categoría de leche. Se analizan 6 productos (todos con formato de 1 litro), que se corresponden con: Asturiana, Ato, MMDD, Pascual Calcio, Pascual Clásica y Puleva Calcio. El comportamiento de un tipo de leche determinado, leche con Calcio, se estudia de manera particular.

La Figura 4.3 muestra la evolución temporal (43 semanas) del índice de precios y de las unidades vendidas para cada producto de la categoría leche. El índice de precios está expresado en tanto por uno: un valor de 1 indica que no se ha realizado ninguna promoción (el producto mantiene su precio regular); un valor inferior, por ejemplo 0.96, indica un descuento del 4% respecto a su precio regular. En la 4.3(a) se muestra la evolución del índice de precios y unidades vendidas del producto Asturiana. En concreto, la leche Asturiana presenta 7 valores distintos de índice de precios : 0.93, 0.94, 0.95, 0.96, 0.97 y 1. De un análisis temporal, se observa que leche Asturiana tiene un volumen de ventas elevado, entorno a 4000 unidades vendidas, alcanzando su máximo durante la sexta semana con un índice de precio de 0.97 (existe descuento). Durante la semana 14 se realiza un descuento directo del 7% que supone un ligero aumento de ventas en la semana 16. No se realiza ningún tipo de promoción entre las semanas 17 y 22, lo cual provoca un descenso de unidades vendidas, que aumentarán en la semana 24 como consecuencia de la bajada de precios realizada en la semana 23. Se puede observar que en la semana 34 el número de unidades vendidas aumenta notablemente, aunque el producto no está en promoción. Por tanto, se puede concluir que la leche Asturiana no es un producto que responda de forma clara al descuento directo, y por tanto su comportamiento se debería estudiar junto con factores más relacionados con el consumidor, como la fidelización a la marca o la percepción real del propio descuento. En la Figura 4.4(a) se muestra su curva de respuesta promocional. Para obtener esta curva, se suma para cada índice de precios el total de unidades vendidas y se divide por el número de semanas con ese índice de precios.

Las unidades vendidas y el índice de precios del producto leche de la marca Ato se representan en la Figura 4.3(b). Se observa que, en términos globales, el número de unidades vendidas es inferior al de Asturiana. No obstante, y como se muestra en la Figura 4.4(b), el volumen de ventas aumenta cuando el producto se encuentra en promoción. La curva de respuesta promocional es lineal decreciente, ya que número de unidades vendidas disminuye a medida que el índice de precios aumenta (reducción del descuento relativo disminuye).

En la Figura 4.3(d) se muestran las unidades vendidas y el índice de precios para Pascual Calcio. Este producto se caracteriza por no tener un volumen de ventas elevado debido a que cubre las necesidades de un segmento de mercado determinado. Cabe destacar que el descuento aplicado en la semana 23 repercute en un aumento considerable de ventas. En la Figura 4.4(d) se presenta la curva de respuesta promocional.

Para leche Pascual Clásica, representada en la Figura 4.3(e), destacan dos efectos. El primero se corresponde con el aumento de ventas experimentado en la semana 30 como consecuencia de la promoción aplicada. El segundo efecto responde a la propia estacionalidad de las ventas de este producto y es objeto de estudio en la literatura: se produce una reducción considerable del volumen de ventas alcanzando valores inferiores a los que solía tener antes de la promoción.

El efecto de estacionalidad también se puede observar en Puleva Calcio (véase Figura 4.3(f)). El descuento de precio aplicado durante las semanas 28, 29 y 30 provoca, por una parte, que en la semana 30 se obtenga el máximo número de unidades vendidas durante el año; por otra, que el número de unidades vendidas posteriormente descienda considerablemente y alcance valores inferiores a los que tenía antes de la promoción. Este hecho también se puede observar en la Figura 4.4(f), donde el número de unidades vendidas disminuye aunque el descuento de precios sea mayor (promoción realizada en la semana 33).

De la comparación de Pascual Calcio y Puleva Calcio, se observa que el número de unidades vendidas de Puleva Calcio es mayor. No obstante, ambos productos reducen considerablemente sus ventas las semanas posteriores a las de una promoción con descuento de precio de 7 y 6 puntos, respectivamente. Este efecto es mucho más acusado en Puleva Calcio, haciendo que su volumen de ventas sea similar al de Pascual Calcio.

Respecto a la marca del distribuidor, como se indica en la Figura 4.3(c) se trata del producto con menos promociones. Cabe destacar el descuento de precio realizado la semana 35, que supone un aumento de ventas en la semana 36, a partir de la cual comienzan a descender. Como se observa en la Figura 4.4(c), el descuento de 15 puntos realizado al final del año no supone un incremento de ventas.

Para explicar algunos comportamientos observados en las figuras anteriores, sería necesario analizar, entre otros:

- Efectos cruzados de las promociones sobre las ventas de otras marcas o categorías de producto.
- Efectos de interacción entre promociones simultáneas dentro de la categoría.
- Efectos de interacción entre diferentes actividades y soportes promocionales.

Categoría de cerveza. Los productos de la categoría de cerveza se caracterizan por un volumen de ventas considerablemente menor al de los productos de la categoría de leche, ya que se dirigen a un segmento de mercado más específico. En esta categoría se analizan 14 productos: Amstel 25 por 6, Amstel 33, Bavaria 33, Cruzcampo, Estrella 25 por 12, Estrella 25 por 6, Estrella 33, Heineken 25 por 6, MMDD 33, San Miguel 25 por 6, San Miguel 25 por 12, Voll Damm 25 por 6, Xibeca 25 por 6 y Xibeca 33.

Se analiza de manera particular el comportamiento según el tipo de formato: 33 centilitros, 6 botellines de 25 centilitros y 12 botellines de 25 centilitros.

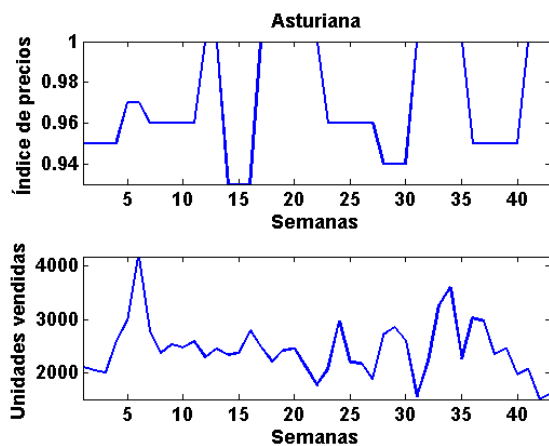
Al realizar un análisis individual de cada producto se observa que Amstel 25 por 6, Cruzcampo, Heineken 25 por 6, MMDD 33, San Miguel 25 por 12 y Xibeca 33 aumentan sus ventas cuando existe un descuento de precio (véanse las Figuras 4.5(a), 4.5(d), 4.6(b), 4.6(c), 4.6(e) y 4.7(b)). Las correspondientes curvas de respuesta promocional, linealmente decrecientes, se muestran en las Figuras 4.8(a), 4.8(d), 4.9(d), 4.9(e), 4.9(f), 4.10(a) y 4.10(d). Respecto a Estrella 33 se observa que no está en promoción durante el año analizado y, sin embargo, el número de unidades vendidas aumenta a finales de año (Figura 4.6(a)).

Las ventas de otros productos como Bavaria 33, Estrella 25 por 6 y Xibeca 25 por 6, disminuyen cuando se realiza un descuento sobre su precio de referencia, como se observa en las Figuras 4.5(c), 4.5(f) y 4.7(a), y en las correspondientes curvas de respuesta promocional (lineal creciente) en las Figuras 4.8(c), 4.9(b) y 4.10(c).

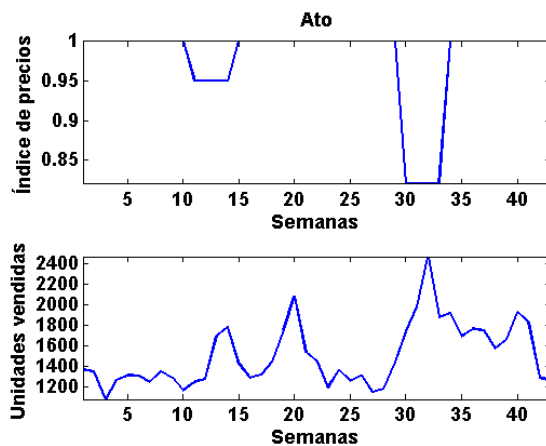
Analizando el formato del producto cerveza, existen 4 marcas con diferentes formatos: Amstel, Estrella, San Miguel y Xibeca. En términos de unidades vendidas, para la marca Amstel es preferible el formato de 33 centilitros, sin embargo el *pack* de 6 botellines de 25 centilitros responde mejor a las promociones realizadas. Para la marca Estrella, es preferible el formato de 33 centilitros en términos de volumen de ventas; en términos de promociones es preferible el *pack* de 12 botellines de 25 centilitros. Para la marca San Miguel, se promociona más el *pack* de 6 botellines de 25 centilitros, aunque no existen diferencias significativas respecto al *pack* de 12 botellines de 25 centilitros. En cuanto a los productos de la marca Xibeca en términos de volumen de ventas es preferible un formato de 33 centilitros tanto cuando existe como cuando no existe promoción.

Un análisis del formato de los productos de la categoría de cerveza sin tener en cuenta la marca indica que, en términos de unidades vendidas, es preferible el formato de 33 centilitros. No obstante, tanto las unidades vendidas de este formato como el *pack* de 6 botellines de 25 centilitros aumentan cuando se realiza un descuento de precio.

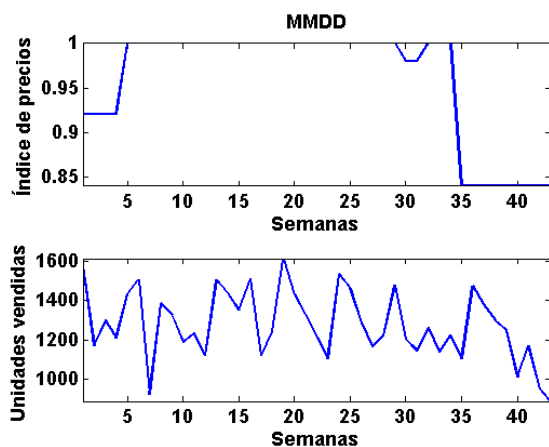
Por último, respecto al comportamiento de las ventas del producto de la marca del distribuidor, se observa que responden de forma directa al descuento de precios realizado.



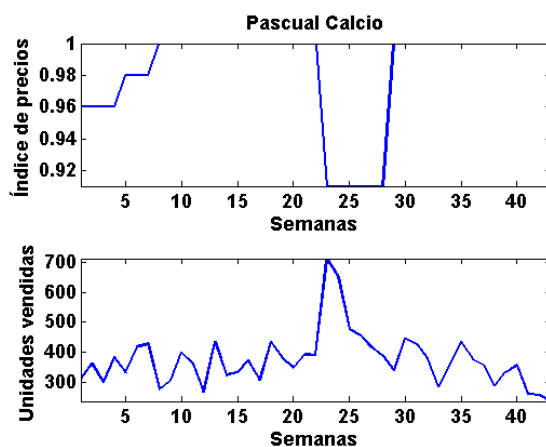
(a)



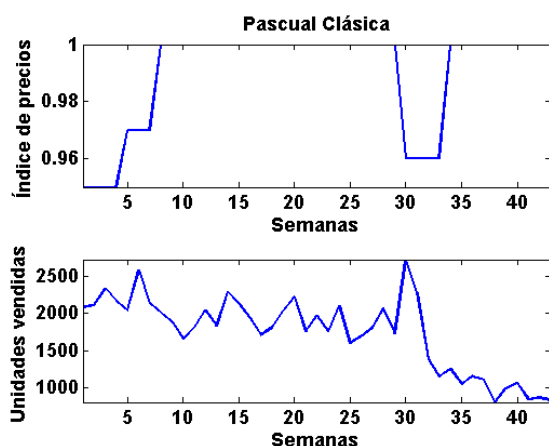
(b)



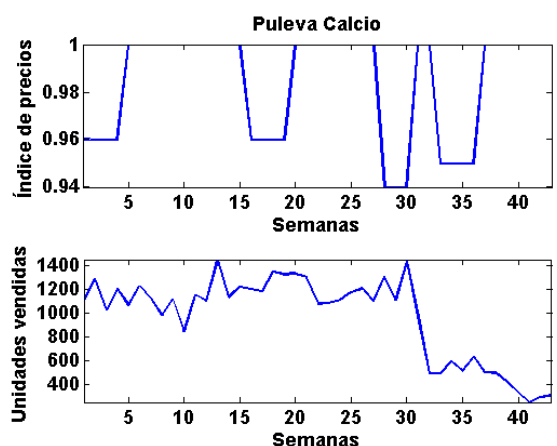
(c)



(d)



(e)



(f)

Figura 4.3: Índice de precios y unidades vendidas de 6 productos de la categoría leche: (a) Asturiana; (b) Ato (c) MMDD; (d) Pascual Calcio; (e) Pascual Clásica; (f) Puleva Calcio.

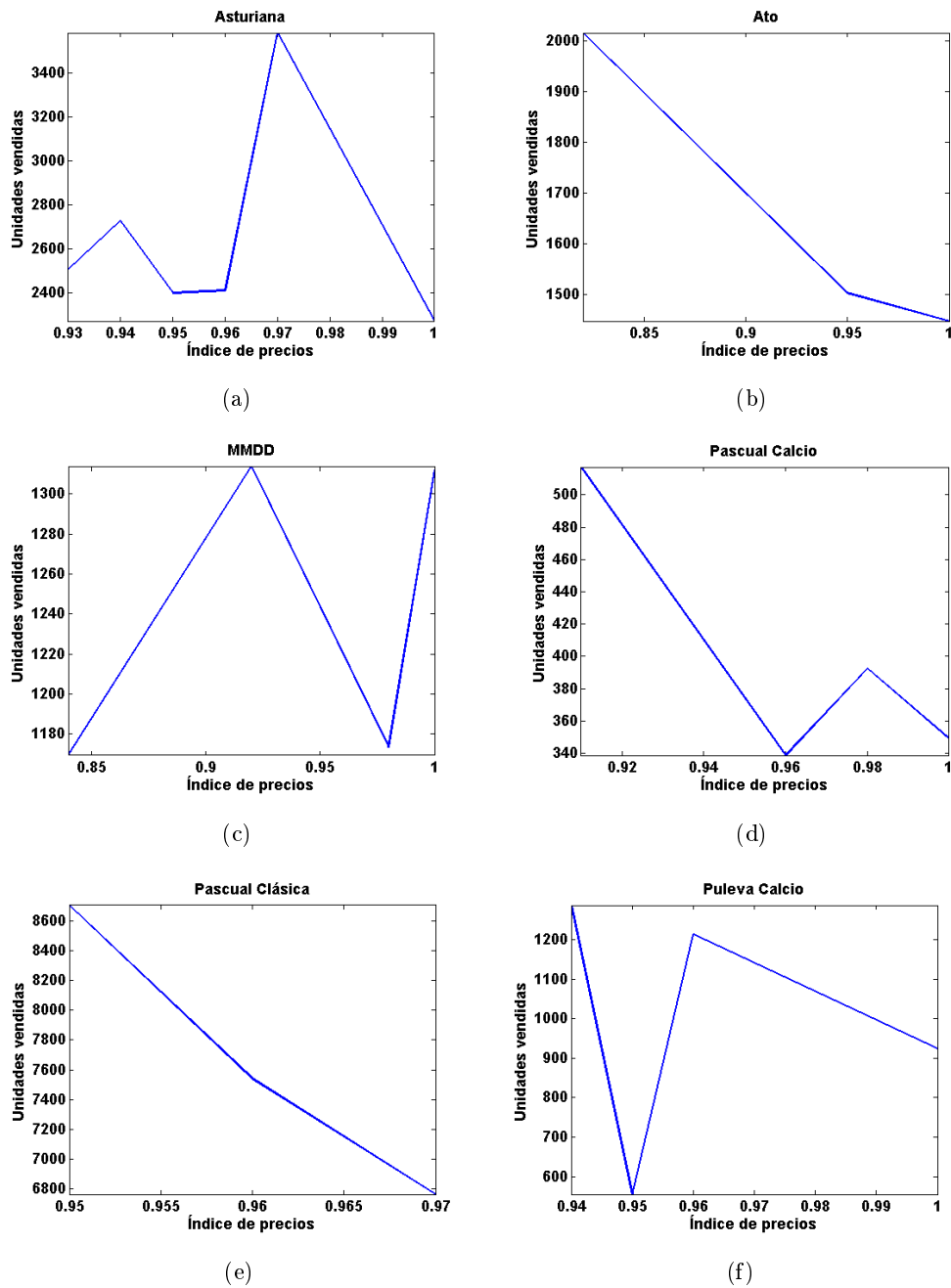


Figura 4.4: Índice de precios vs unidades vendidas de 6 productos de la categoría de leche: (a) Asturiana; (b) Ato (c) MMDD; (d) Pascual Calcio; (e) Pascual Clásica; (f) Puleva Calcio.

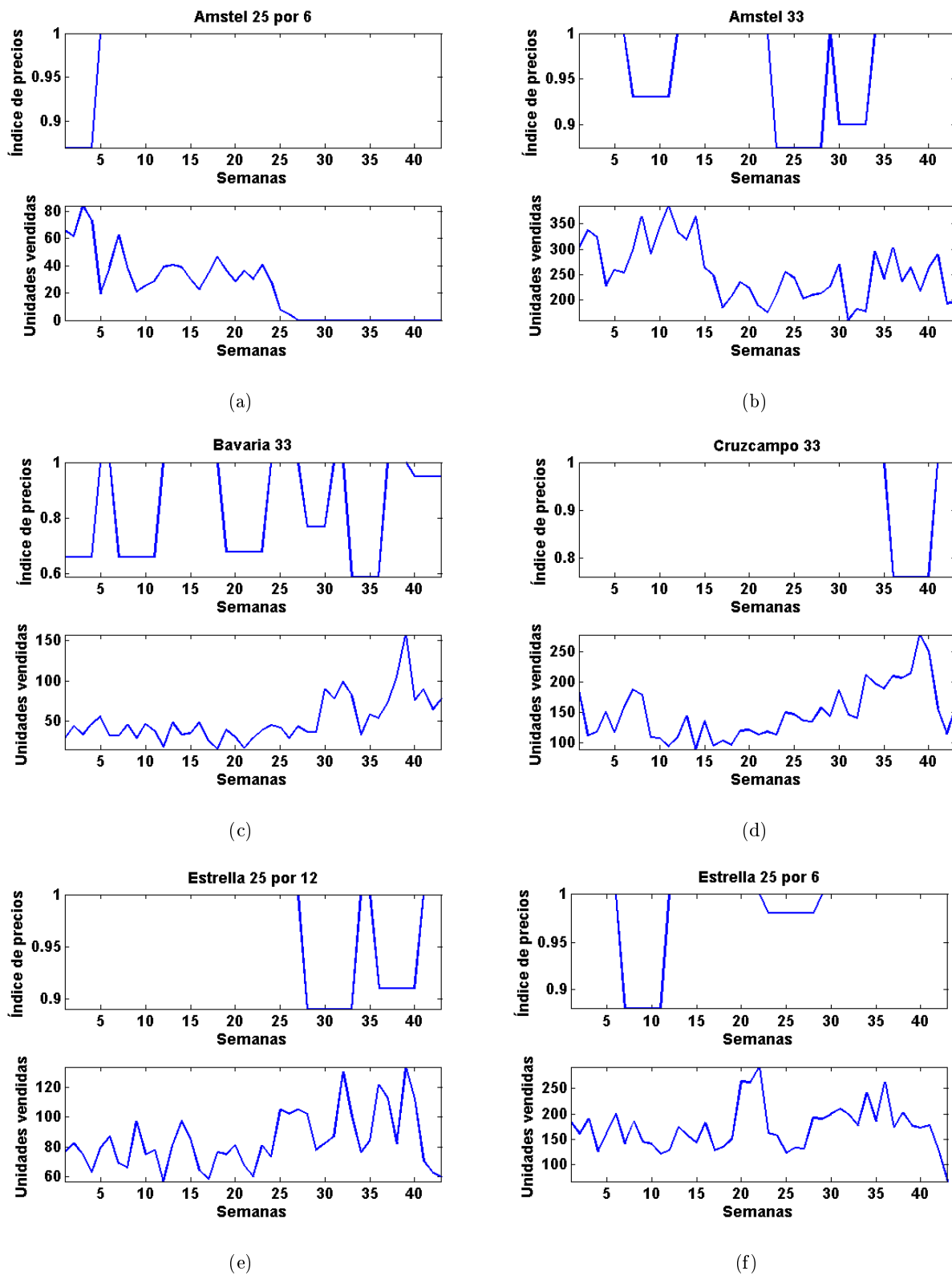


Figura 4.5: Índice de precios y unidades vendidas de 6 productos de la categoría de cerveza: (a) Amstel 25 por 6; (b) Amstel 33; (c) Bavaria 33; (d) Cruzcampo; (e) Estrella 25 por 12; (f) Estrella 25 por 6.

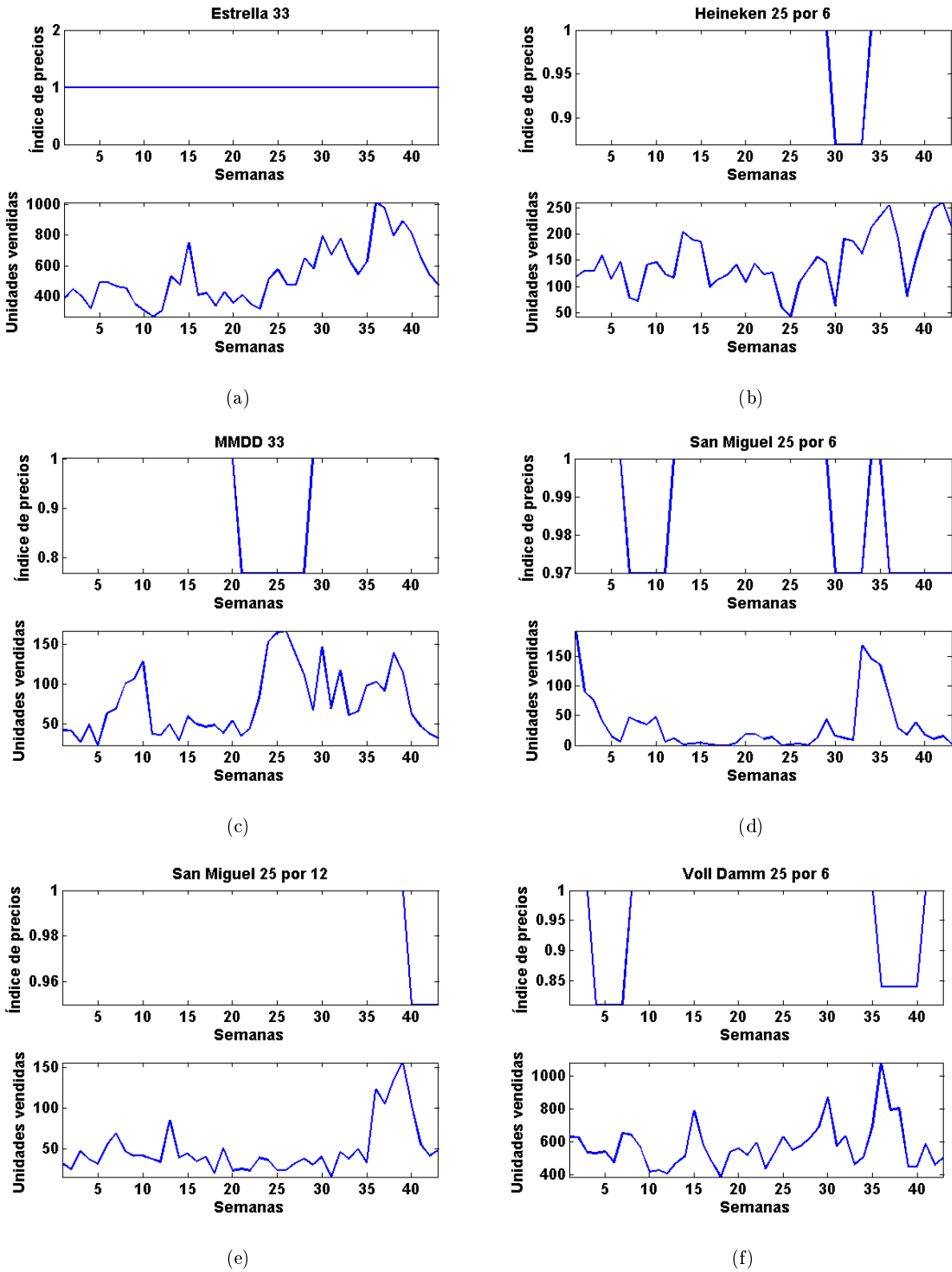


Figura 4.6: Índice de precios y unidades vendidas de 6 productos de la categoría de cerveza: (a) Estrella 33; (b) Heineken 25 por 6; (c) MMDD 33; (d) San Miguel 25 por 6 (e) San Miguel 25 por 12; (f) Voll Damm 25 por 6.

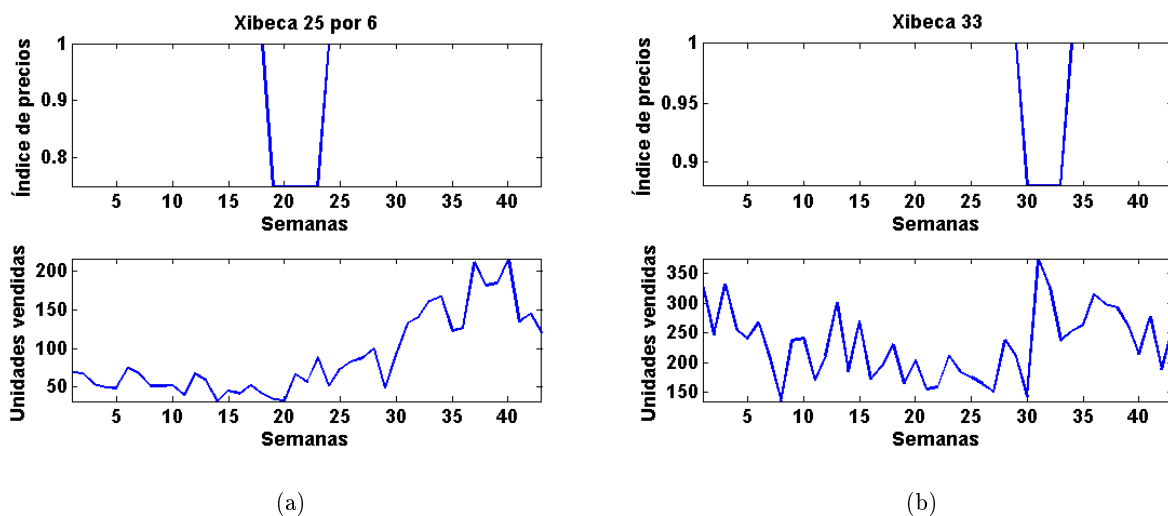


Figura 4.7: Índice de precios y unidades vendidas de dos productos de la categoría de cerveza: (a) Xibeca 25 por 6; (b) Xibeca 33.

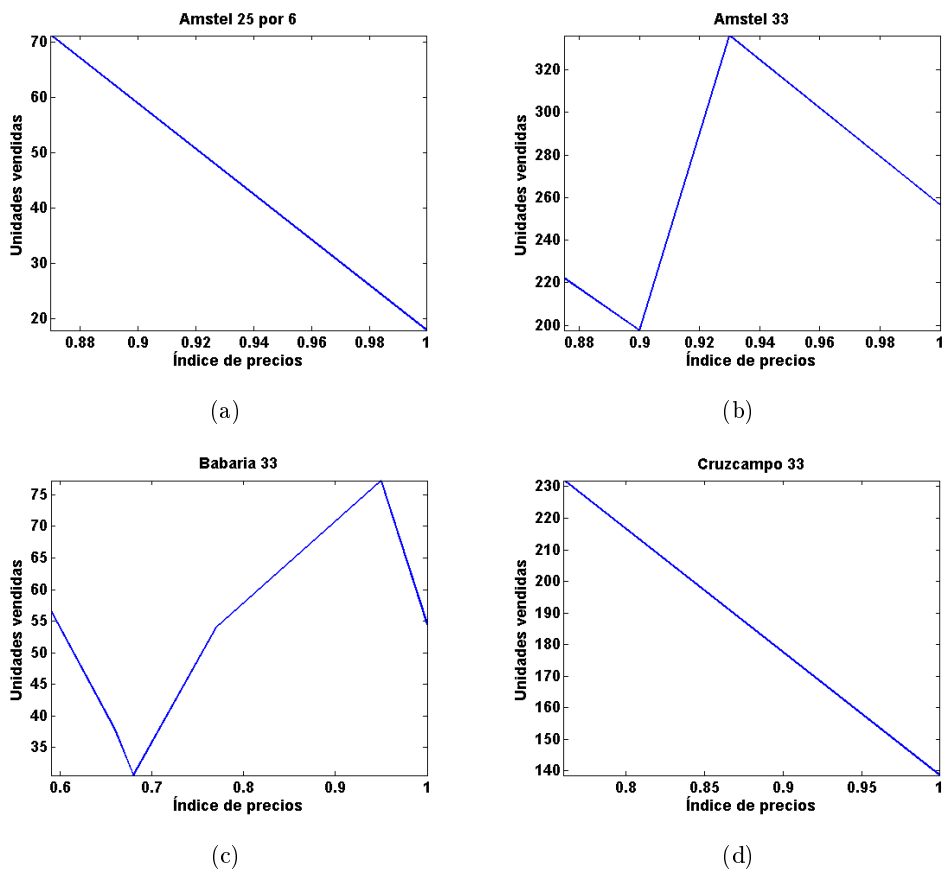


Figura 4.8: Índice de precios vs unidades vendidas de cuatro productos de cerveza: (a) Amstel 25 por 6; (b) Amstel 33 (c) Bavaria 33; (d) Cruzcampo.

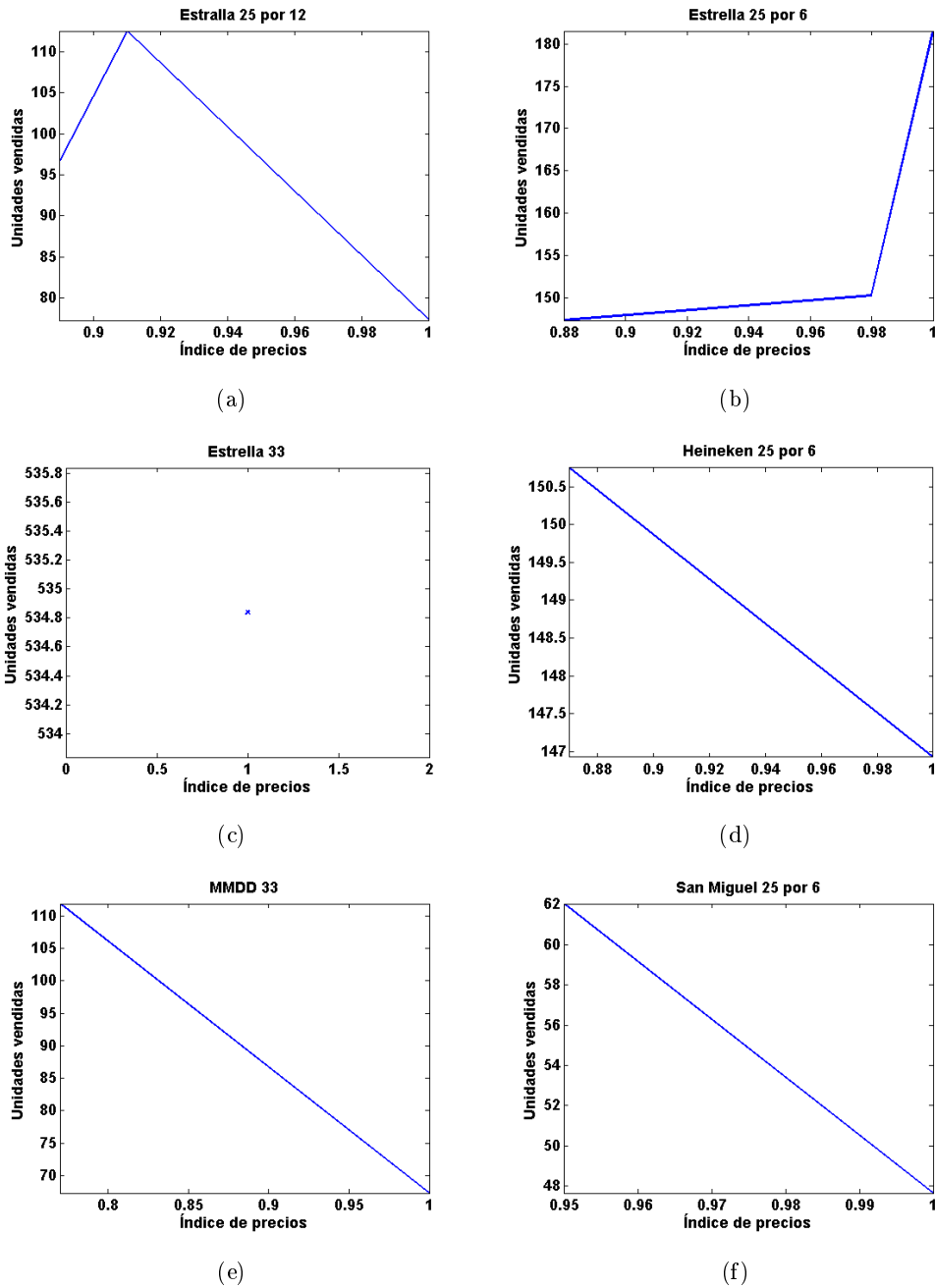


Figura 4.9: Índice de precios vs unidades vendidas de 6 productos de la categoría de cerveza: (a) Estrella 25 por 6; (b) Estrella 25 por 12; (c) Estrella 33; (d) Heineken 25 por 6; (e) MMDD 33; (f) San Miguel 25 por 6.

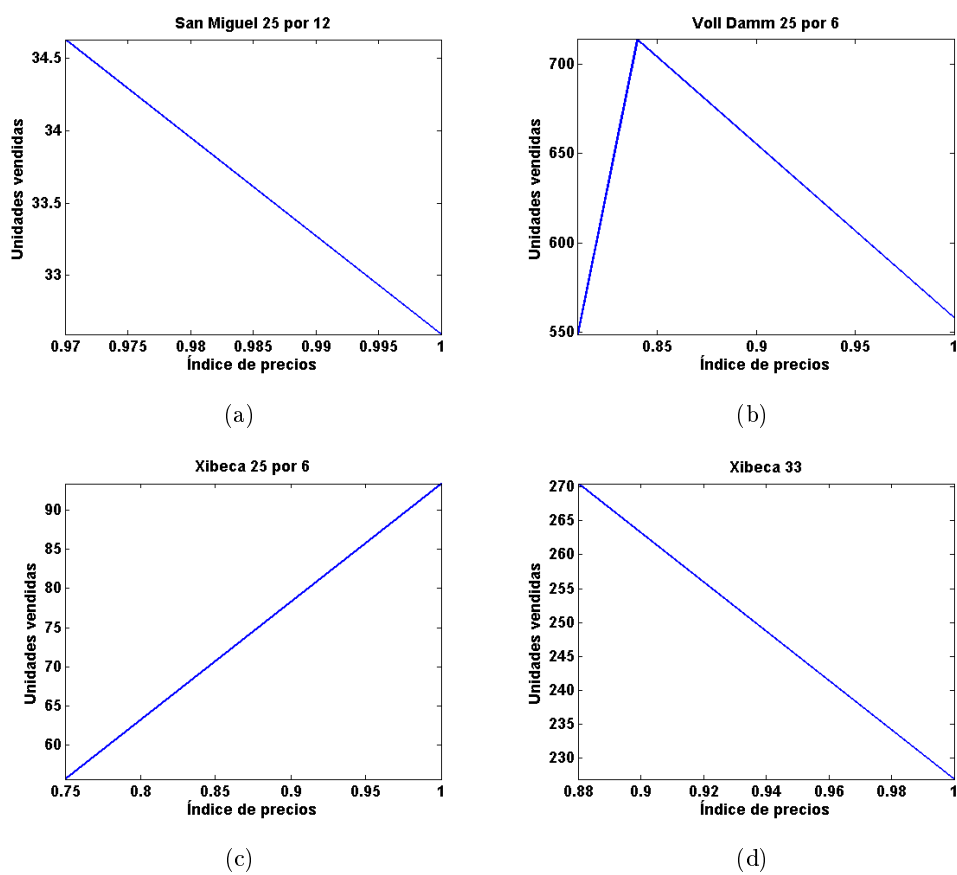


Figura 4.10: Índice de precios vs unidades vendidas de cuatro productos de la categoría de cerveza: (a) San Miguel 25 por 12; (b) Voll Damm 25 por 6; (c) Xibeca 25 por 6; (d) Xibeca 33.

Análisis y resultados con datos minoristas

5.1. Introducción

En este capítulo se analizan dos bases de datos reales de las categorías de productos de leche y cerveza, ambas procedentes de un distribuidor minorista. El objetivo es comprobar qué métodos de aprendizaje estadístico tienen mejor comportamiento para analizar las respuestas promocionales. Existen en la literatura estudios similares con la categoría de café [22] [25]. Estos estudios hacen uso de un esquema ν -SVM para diseñar el modelo, y realizan comparaciones con métodos estadísticos clásicos (como la regresión lineal o la regresión semiparamétrica), aunque no con otros esquemas basados en aprendizaje estadístico. Dichos estudios tampoco realizan ningún test estadístico para determinar qué algoritmo clásico ofrece mejores prestaciones ante una base de datos dada (tarea, por otro lado, no trivial).

En este proyecto se han construido modelos de promociones de ventas utilizando los cuatro esquemas de aprendizaje descritos en el Capítulo 3 (k-NN, GRNN, MLP y ν -SVM). Para seleccionar los parámetros libres de cada modelo se ha utilizado la técnica de validación cruzada LOO, por disponer de un número de ejemplos muy reducido (43). Para todos los experimentos, la figura de mérito considerada en LOO ha sido el MAE. Así, se obtienen las prestaciones de cada modelo en función del parámetro libre y se elige aquel modelo con mejores prestaciones.

No obstante, y como se presentará más adelante, es complicado decidir qué método es el mejor simplemente mediante la comparación de una única figura de mérito (MAE). Por otra parte, la utilización de tests estadísticos convencionales puede ser, o bien cuestionable (debido a las distribuciones no gaussianas de los errores), o bien tan compleja de manejar que se torne no operativa. Por ello, en este trabajo se realiza un diagnóstico del modelo utilizando remuestreo *bootstrap*, técnica estadística no paramétrica que permite establecer criterios de decisión sobre la existencia de diferencias significativas entre las prestaciones de dos modelos.

La estructura del resto del capítulo es la siguiente. En la Sección 5.2 se explica y realiza la

búsqueda de parámetros libres en los métodos de aprendizaje estadístico utilizados. En el resto de secciones se presentan 5 experimentos: el Experimento 1 compara diferentes métodos en términos de MAE; el Experimento 2 compara los métodos usando un test no paramétrico que simultáneamente tiene en cuenta medidas de centrado y dispersión del MAE; en el Experimento 3 se comparan las prestaciones del esquema ν -SVM diseñado con núcleo RBF y con núcleo semi-paramétrico; en el Experimento 4 se profundiza en los elementos de diseño del MLP; por último, en el Experimento 5 se analiza la conveniencia de incluir determinadas variables (en general de naturaleza dicotómica) en los modelos promocionales.

5.2. Elección de parámetros libres

Para elegir los parámetros libres se ha considerado el MAE como figura de mérito y se ha aplicado la técnica de validación LOO. Puesto que el procedimiento es el mismo para las dos categorías de productos consideradas (leche y cerveza), sólo se presentan de forma gráfica los resultados obtenidos para la categoría de leche.

Categoría de leche

k-NN. Como se ha explicado en el Capítulo 3, k-NN es un método no paramétrico cuya salida depende del parámetro libre k (número de vecinos considerado). Valores bajos de k hacen que el modelo sea muy sensible a muestras atípicas; valores altos hacen que la estima pueda diferir demasiado del valor real debido al sobrepromediado.

Para cada modelo de la categoría de leche se han explorado valores de k en el intervalo [1,40]. En la Figura 5.1 se muestra la evolución del MAE sobre el número de unidades vendidas semanalmente, aplicando la técnica LOO para cada modelo y explorando distintos valores de k considerados. Aunque se ha marcado el mínimo valor de MAE, se podría considerar cualquier valor de k en el rango en el que el MAE se mantiene más o menos estable.

GRNN. Esta red proporciona múltiples salidas, por lo que se obtiene una única función de regresión para estimar de forma conjunta las unidades vendidas de cada producto. No obstante, hay que señalar que las salidas están en realidad desacopladas y no modelan interacciones entre las mismas. Considerando funciones gaussianas esféricas, la GRNN tiene un único parámetro libre σ (anchura de las funciones base). En la Figura 5.2 se muestra el intervalo de exploración considerado [0.01,0.05] y el valor que minimiza el MAE aplicando LOO.

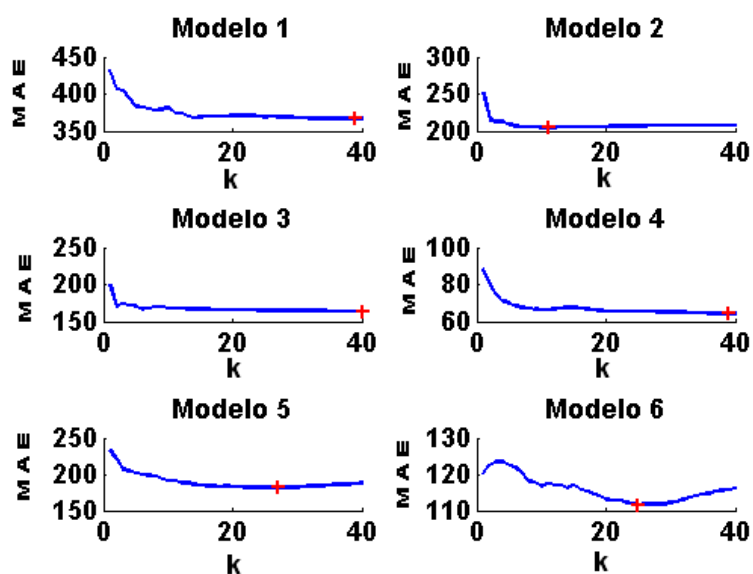


Figura 5.1: Categoría de leche. Evolución del MAE con k para k -NN, indicándose en rojo el valor de k que obtiene el mínimo MAE.

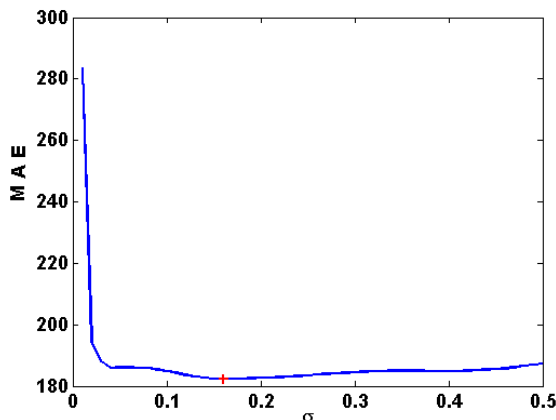


Figura 5.2: Categoría de leche. Evolución del MAE con σ para GRNN. Se indica en rojo el valor de σ que proporciona el mínimo MAE.

MLP. Esta red también se ha implementado con salidas múltiples, pero en este caso las salidas no están desacopladas. Puesto que se ha considerado una red con una única capa oculta, el parámetro libre es el número de neuronas ocultas n_o . Como se indicó en el Capítulo 3, para determinar los pesos y umbrales del MLP se ha utilizado el algoritmo iterativo Levenberg-Marquardt junto con un procedimiento de *early-stopping*. Este procedimiento evita sobreaprendizaje dividiendo en dos subconjuntos los ejemplos utilizados en el diseño:

- Subconjunto de entrenamiento propiamente dicho, utilizado para ajustar pesos y umbrales.
- Subconjunto de validación, para evaluar la respuesta de la red ante patrones no utilizados en el subconjunto de entrenamiento.

El procedimiento *early-stopping* detiene el aprendizaje cuando las prestaciones sobre el conjunto de validación empeoran o se mantienen estables. En la Figura 5.3 se muestra la evolución del MAE (aplicando LOO) con el número de neuronas de la capa oculta, indicando el valor de n_o para el que se obtiene el mínimo MAE.

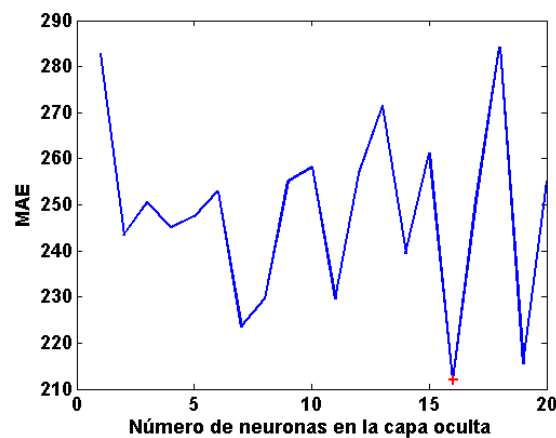


Figura 5.3: Categoría de leche. Evolución del MAE con el número de neuronas de la capa oculta para MLP. Se indica en rojo el valor de n_o que obtiene el mínimo MAE.

SVM. Para realizar el diseño del modelo promocional con SVM se han considerado dos esquemas: SVM convencional y ν -SVM. La utilización de SVM convencional implica la búsqueda de dos parámetros libres: la anchura σ del núcleo de la RBF y el parámetro de regularización C . Los resultados obtenidos con este esquema no han sido de calidad suficiente debido a las características de las bases de datos utilizadas, por lo que no se presentan en esta memoria.

El segundo esquema de tipo SVM utilizado para estimar las unidades vendidas ha sido ν -SVM, con tres parámetros libres: C , σ y ν , donde ν es el límite inferior de la fracción de vectores soporte. Para determinar el valor de los tres parámetros libres sin realizar una exploración en un espacio de tres dimensiones, lo que conllevaría un elevado esfuerzo computacional, se ha utilizado un procedimiento iterativo que parte de un valor inicial de C conforme a la siguiente metodología:

1. Fijado C , se realiza un barrido en el espacio (ν, σ) (véase Figura 5.4(a)).
2. Se elige el par (ν, σ) que minimiza una figura de mérito: MAE para los experimentos de este capítulo, y MSE y D para los mostrados en el Apéndice A.

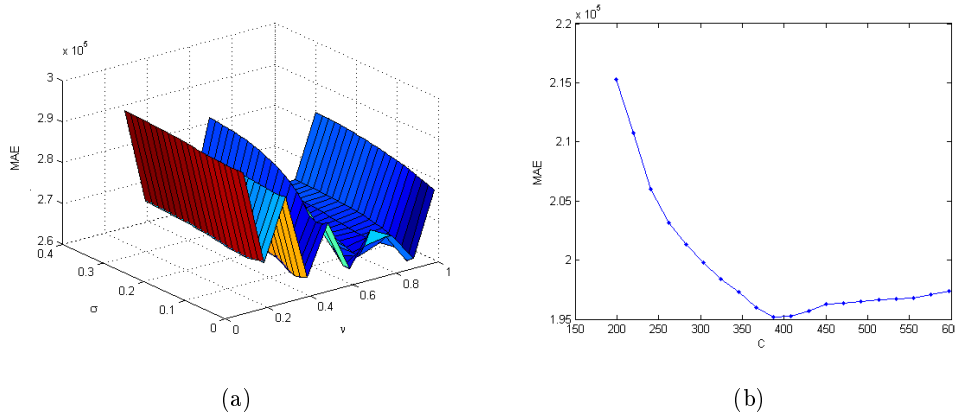


Figura 5.4: *Categoría de leche. Evolución del MAE con: (a) $\nu - \sigma$ con C fijo; (b) C con (σ, ν) fijo.*

3. Con los valores de (ν, σ) obtenidos en el paso 2, se realiza un barrido del parámetro de regularización C , y se elige aquel valor que mejores prestaciones ofrece (véase Figura 5.4(b)).
4. Se calcula el MAE resultante y se vuelve al paso 1. El proceso de búsqueda se detiene cuando el MAE se estabiliza.

El procedimiento anterior se ha realizado en este proyecto con cuatro inicializaciones de $C = 10, 50, 100$ y 1000 para cada uno de los modelos promocionales.

En el artículo [25] sobre modelado promocional con técnicas SVM, los autores presentan una comparativa entre regresión semiparamétrica convencional y basada en SVM. Para diseñar el esquema SVM semiparamétrico, es necesario realizar la búsqueda de un parámetro libre adicional (parámetro de compensación to), correspondiente al núcleo semiparamétrico. El procedimiento seguido en este proyecto para la elección de los 4 parámetros libres es el descrito para ν -SVM, pero incluyendo una etapa adicional de búsqueda del parámetro de compensación una vez fijada la tupla (C, ν, σ) . En el Experimento 3 se detallan los resultados obtenidos al comparar los esquemas ν -SVM y SVM semiparamétrico.

En la Tabla 5.1 se muestra, para cada modelo (producto), el valor de un parámetro libre obtenido por LOO.

Categoría de cerveza

En la Tabla 5.2 se muestran los parámetros libres de los métodos k-NN, GRNN, MLP y ν -SVM para los productos de la categoría de cerveza. Para los métodos con entradas y salidas múltiples (GRNN y MLP), el parámetro libre de la GRNN es el mismo que el que se obtiene con los ejemplos de la categoría de leche; la red neuronal precisa de una neurona más (17 vs 16) para poder estimar las unidades vendidas de la categoría de cerveza.

	k (k-NN)	σ (GRNN)	n_o (MLP)	σ (ν -SVM)
Modelo 1	39	0.16	16	0.02
Modelo 2	11	0.16	16	0.22
Modelo 3	40	0.16	16	0.51
Modelo 4	39	0.16	16	0.03
Modelo 5	27	0.16	16	0.43
Modelo 6	25	0.16	16	0.49

Tabla 5.1: *Categoría de leche. Parámetros libres seleccionados aplicando LOO sobre cada método y modelo.*

	k (k-NN)	σ (GRNN)	n_o (MLP)	σ (ν -SVM)
Modelo 1	3	0.16	8	0.28
Modelo 2	40	0.16	8	0.28
Modelo 3	11	0.16	8	0.28
Modelo 4	8	0.16	8	0.01
Modelo 5	3	0.16	8	0.39
Modelo 6	40	0.16	8	0.08
Modelo 7	3	0.16	8	0.18
Modelo 8	12	0.16	8	0.24
Modelo 9	31	0.16	8	0.16
Modelo 10	27	0.16	8	0.20
Modelo 11	38	0.16	8	0.65
Modelo 12	1	0.16	8	0.16
Modelo 13	19	0.16	8	0.21
Modelo 14	37	0.16	8	0.12

Tabla 5.2: *Categoría de cerveza. Valores de los parámetros libres seleccionados al aplicar LOO sobre cada método y modelo.*

5.3. Experimento 1: comparativa de métodos

El objetivo de este primer experimento es determinar qué esquema de aprendizaje (k-NN, GRNN, MLP y ν -SVM) ofrece mejores prestaciones en términos de MAE para cada categoría de producto.

Categoría de leche. El modelo a diseñar considera 8 variables de entrada: 6 variables métricas, correspondientes a los índices de precios de cada producto, y dos variables dicotómicas (indicador de descuento directo e indicador de estacionalidad). El número de salidas del modelo depende del método considerado: 1 salida para k-NN y ν -SVM, y 6 salidas para GRNN y MLP.

Se ha estimado el número de unidades vendidas para cada modelo, y este valor se ha comparado con el número de unidades vendidas reales. Los valores de MAE obtenidos para cada

método se muestran en la Tabla 5.3. Para cada celda de la tabla existen dos filas, el primer valor de la primera fila se corresponde con el MAE empírico, el segundo valor muestra el promedio del MAE usando *bootstrap*; en la segunda fila aparece el IC al 95 % del MAE bootstrapeado. Para cada modelo se indica en negrita el método que mejores prestaciones ofrece. De los valores mostrados en la Tabla 5.3 se deduce que ν -SVM es el mejor esquema para los Modelos 2, 3 y 4, y k-NN es el mejor para el Modelo 5. Para los Modelos 1 y 6, ν -SVM ofrece prestaciones similares a MLP y GRNN, respectivamente.

	k-NN	GRNN	MLP	ν -SVM
Modelo 1	366.44 365.50 [269.95,465.16]	412.42 410.61 [314.41,532.60]	320.75 320.56 [247.23,407.72]	322.15 320.68 [231.08,422.75]
Modelo 2	205.21 205.56 [156.94,257.97]	187.57 187.59 [139.29,235.98]	198.89 198.22 [154.57,244.11]	166.09 166.13 [123.03,214.02]
Modelo 3	164.62 165.11 [131.26,199.64]	150.77 150.68 [125.99,177.73]	151.23 151.83 [122.15,182.68]	136.12 135.86 [109.63,165.34]
Modelo 4	63.27 63.18 [48.61,79.27]	63.89 63.53 [47.76,83.19]	72.41 72.85 [58.06,89.09]	50.64 50.41 [37.87,65.05]
Modelo 5	178.99 178.49 [131.93,226.97]	193.34 194.21 [154.14,238.14]	196.36 196.83 [151.58,248.20]	192.59 193.03 [134.98,257.77]
Modelo 6	119.68 119.54 [92.13,148.40]	104.42 104.38 [78.14,131.18]	127.68 127.64 [99.38,157.79]	105.13 105.24 [80.24,132.32]

Tabla 5.3: *Categoría de leche. Primera fila de cada celda: MAE empírico y MAE promedio. Segunda fila de cada celda: IC al 95 %. En negrita se indica el método con mejores prestaciones para cada producto.*

Categoría de cerveza. El modelo a diseñar considera 16 variables de entrada: 14 variables métricas, correspondientes a los índices de precios de cada producto, y dos variables dicotómicas (indicador de descuento directo e indicador de estacionalidad). El número de salidas depende del método considerado: 1 salida para k-NN y ν -SVM, y 14 salidas para GRNN y MLP.

Los resultados obtenidos en términos de MAE se muestran en la Tabla 5.4. Obsérvese que ν -SVM es el método que mejores prestaciones ofrece salvo en los productos 2 y 7, para los que k-NN es el mejor esquema.

Como conclusiones de este primer experimento, de forma general, se puede afirmar que usando como figura de mérito MAE en términos absolutos, ν -SVM es el método que proporciona mejores prestaciones, siendo este comportamiento más notable en los productos de cerveza. Ocasionalmente, k-NN ofrece mejores prestaciones que ν -SVM, al menos en términos de MAE.

Respecto al resto de métodos (k-NN, GRNN y MLP), es complicado decidir qué método es mejor en términos de MAE. Por ejemplo, para la categoría leche, GRNN presenta menor MAE

	k-NN	GRNN	MLP	ν -SVM
Modelo 1	7.85 7.85 [5.16,10.94]	11.85 11.90 [9.07,14.83]	13.86 13.75 [10.33,17.50]	6.98 6.96 [4.77,9.16]
Modelo 2	33.77 33.61 [25.67,42.11]	47.56 47.43 [38.10,57.32]	49.76 49.42 [39.08,62.28]	34.52 34.53 [27.33,42.32]
Modelo 3	15.10 14.96 [11.48,19.38]	14.92 14.70 [10.94,19.35]	14.33 14.28 [10.49,18.60]	14.08 14.11 [10.48,18.67]
Modelo 4	25.75 25.74 [19.59,32.09]	27.51 27.58 [22.34,33.32]	26.74 26.42 [20.15,34.00]	21.40 21.40 [16.36,27.01]
Modelo 5	12.21 12.05 [8.82,15.92]	13.67 13.67 [10.67,16.49]	13.62 13.54 [10.28,16.77]	10.27 10.25 [7.54,13.19]
Modelo 6	31.10 30.89 [23.82,39.15]	32.33 32.15 [24.29,40.91]	31.59 31.65 [25.15,38.95]	26.82 26.86 [18.92,35.37]
Modelo 7	78.86 78.57 [61.13,99.64]	97.54 96.86 [72.26,123.36]	103.53 102.14 [72.98,136.63]	89.50 89.53 [64.13,118.31]
Modelo 8	36.90 36.92 [29.91,48.91]	36.19 36.19 [26.89,46.31]	47.42 47.15 [38.34,57.09]	34.59 34.71 [26.83,43.39]
Modelo 9	20.77 20.64 [16.04,25.67]	32.93 32.85 [26.50,39.82]	28.92 28.70 [21.18,38.62]	19.58 19.62 [14.48,25.02]
Modelo 10	17.98 17.88 [11.54,26.99]	24.12 24.02 [16.44,34.14]	26.08 26.22 [18.29,35.32]	12.85 12.84 [8.39,19.15]
Modelo 11	13.65 13.62 [9.74,17.73]	17.48 17.42 [12.51,23.41]	13.60 13.62 [9.86,18.06]	11.48 11.43 [7.92,15.18]
Modelo 12	109.58 109.69 [83.88,140.81]	102.62 102.00 [77.19,130.96]	146.30 146.07 [107.64,183.87]	96.85 97.06 [71.72,126.61]
Modelo 13	16.49 16.45 [12.88,20.21]	22.83 22.69 [18.21,27.91]	26.16 25.92 [18.84,34.56]	14.91 14.94 [11.47,18.78]

Tabla 5.4: *Categoría de cerveza. Primera fila de cada celda: MAE empírico y MAE promedio. Segunda fila de cada celda: IC al 95%. En negrita se indica el método con mejores prestaciones para cada producto.*

que k-NN y MLP para los productos 2 y 3, pero no para el resto de productos. Por ello, y para además verificar que las diferencias en las prestaciones de los métodos en términos de MAE de las ν -SVM son realmente significativas, se propone usar un método de remuestreo *bootstrap* no paramétrico que permita contrastar las conclusiones extraídas estableciendo un test de hipótesis.

5.4. Experimento 2: comparativa mediante *bootstrap* pareado

Como se ha visto en los resultados presentados en el primer experimento, resulta complicado decidir qué métodos de los analizados presentan mejores prestaciones simplemente mediante

inspección de las figuras de mérito, dado que los valores en ocasiones son ligeramente diferentes.

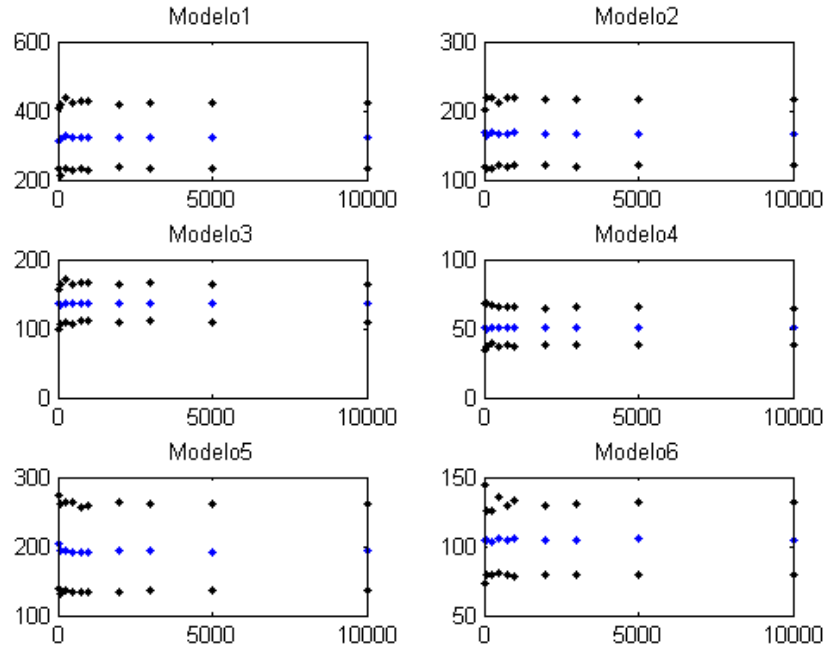


Figura 5.5: *MAE* obtenido al realizar *bootstrap* con diferente número de remuestras para los modelos de la categoría de leche.

Por ello, para contrastar en términos estadísticos con un nivel de confianza del 95 % las diferencias existentes en los resultados anteriores, se utiliza un método no paramétrico denominado remuestreo *bootstrap* pareado, y como se ha visto en la Sección 3.5 tiene como parámetro libre el número de remuestreos B . Como un paso previo, se ha de decidir cuál es el valor de B , y para ello, se ha probado el método *bootstrap* con B igual a 50, 100, 250, 500, 750, 1000, 2000, 3000, 5000 y 10000 remuestreos sobre el intervalo de confianza del MAE considerando ν -SVM como método de aprendizaje estadístico. En la Figura 5.5 se observa cómo para valores superiores a 1000, el intervalo de confianza del MAE obtenido se estabiliza. Por ello, en el resto de este trabajo, se decide tomar $B = 2000$ remuestras para obtener los estadísticos de la media y el intervalo de confianza del MAE sobre las unidades vendidas para caracterizar las prestaciones.

5.4.1. Leche

En la Tabla 5.5 se comparan de forma pareada las prestaciones de k-NN vs GRNN, k-NN vs ν -SVM y GRNN vs ν -SVM, con la finalidad de contrastar si un método es estadísticamente mejor que otro en la categoría de leche. Se realiza un remuestreo *bootstrap* con 3 estadísticos

	k-NN vs GRNN	k-NN vs ν -SVM	GRNN vs ν -SVM
Modelo 1	-33.26 [-107.37,43.21] 150.41 [-453.69,749.42] -54.43 [-499.08,438.29]	45.14 [-6.46,106.76] 231.63 [-344.72,706.90] 8.06 [-247.29,163.40]	76.78 [20.21,133.73] 86.76 [-469.78,431.90] 158.71 [-304.17,300.46]
Modelo 2	25.60 [0.09,54.63] -1.98 [-135.32,121.75] 21.94 [-106.26,194.54]	38.22 [10.13,68.02] 21.14 [-169.14,226.54] 59.74 [-13.31,234.87]	12.13 [1.04,23.53] 18.35 [-21.31,93.62] 29.16 [2.40,109.07]
Modelo 3	16.95 [-1.18,33.74] 141.97 [46.41,246.08] 92.05 [14.20,142.75]	27.40 [7.47,48.29] 159.73 [39.62,260.05] 95.50 [-22.02,180.99]	10.40 [2.19,19.08] 15.92 [-37.84,104.71] 4.50 [-33.02,39.79]
Modelo 4	1.74 [-15.80,14.41] -12.40 [-129.01,144.81] -52.49 [-122.24,98.23]	13.64 [3.43,23.68] 67.71 [-11.58,130.99] 10.79 [-35.07,74.79]	15.49 [4.53,26.47] 82.00 [-13.82,122.43] 64.60 [-32.74,92.66]
Modelo 5	-15.22 [-55.52,28.75] 70.56 [-142.65,255.19] 1.10 [-63.07,57.34]	-11.40 [-52.92,31.54] -41.69 [-417.52,251.53] -202.82 [-401.77,-3.48]	2.86 [-39.87,44.85] -110.44 [-353.20,148.94] -203.46 [-407.88,23.06]
Modelo 6	4.04 [-19.98,27.76] 46.78 [-106.24,163.55] 29.39 [-36.50,111.73]	7.28 [-21.99,36.11] 55.53 [-92.50,169.06] 13.35 [-24.32,60.57]	2.62 [-12.44,17.51] 9.87 [-112.51,104.06] -15.63 [-51.16,38.43]

Tabla 5.5: *Bootstrap pareado para los productos de leche considerando 3 estadísticos diferentes: ΔMAE (primera fila), ΔIC (segunda fila) y ΔIC_{sup} (tercera fila). En cada figura de mérito se indican la media y el intervalo de confianza. En negrita se resaltan los métodos que presentan diferencias estadísticamente significativas con un nivel del 95 %.*

diferentes presentados en el Capítulo 3: la variación del error absoluto medio (ΔMAE , primera fila), la diferencia de la anchura de los intervalos de confianza de cada uno de los dos métodos comparados (ΔIC , segunda fila), y la diferencia del límite superior del intervalo de confianza de cada uno de los métodos (ΔIC_{sup} , tercera fila). Los intervalos de confianza proporcionan una idea de la dispersión del valor obtenido. Se puede decir que 2 métodos son estadísticamente diferentes si el cero no pertenece al intervalo dado. Es decir, si los dos límites del IC tienen signo negativo, el primer método presenta mejores prestaciones; si ambos límites tienen signo positivo, es el segundo método el que presenta mejores prestaciones.

Si se comparan las salidas obtenidas de los métodos k-NN y GRNN, en términos de MAE, solo existen diferencias significativas para el Modelo 2 (véase el Histograma 5.6(a)), y el error obtenido al estimar con GRNN es significativamente menor. Si se analiza el resultado obtenido cuando se considera como estadístico ΔIC , se observa que sólo existen diferencias significativas para el Modelo 3, siendo la anchura del IC significativamente menor cuando se modela con GRNN (véase el Histograma 5.6(b)). Además se observa que para este mismo modelo, el límite superior del IC es mayor cuando se utiliza el esquema k-NN, como se muestra en el Histograma 5.6(c).

Si se comparan los resultados obtenidos de realizar un test no paramétrico para comparar las diferencias existentes entre k-NN y ν -SVM, el MAE obtenido cuando se diseña con ν -SVM es significativamente menor para los Modelos 2, 3 y 4, como se observa en el Histograma 5.7(a). Además, para el Modelo 3, la anchura del IC es significativamente menor para los resultados obtenidos con ν -SVM. En los histogramas de la Figura 5.7(b) se observa que para el Modelo 5, el límite superior del IC es menor cuando se utiliza el esquema k-NN.

De los valores mostrados en la tercera columna de la Tabla 5.5, se puede decir que es preferible estimar el número de unidades vendidas con el método ν -SVM para los productos 1, 2, 3 y 4. En la Figura 5.8(a) se observa que los histogramas correspondientes a estos 4 modelos tienen tanto su media como el valor inferior y superior del IC a la derecha del cero. Sin embargo, en términos de ΔIC , tal y como se muestra en los histogramas de la Figura 5.8(b) no existen diferencias significativas entre GRNN y ν -SVM, y todos los histogramas están centrados en torno al cero. Si se considera ΔIC_{sup} como estadístico, para los datos del Modelo 2, el límite superior del IC obtenido cuando se estima con ν -SVM es menor que cuando se estima con k-NN.

Los Modelos 2 y 3 corresponden a la marca ATO y a la marca del distribuidor respectivamente, entendidas ambas como marca del distribuidor. Para estos dos modelos, las salidas estimadas usando como método de aprendizaje estadístico ν -SVM presenta mejores prestaciones en términos de MAE que usando k-NN o GRNN, si bien no existen diferencias significativas en términos de dispersión.

5.4.2. Cerveza

Para comparar qué método presenta mejores prestaciones cuando se trabaja con datos de las promociones de ventas realizadas en la categoría de cerveza, se ha seguido el mismo procedimiento que el realizado con los datos de la categoría de leche. En la Tabla 5.6 se comparan de forma pareada k-NN vs GRNN, k-NN vs ν -SVM y GRNN vs ν -SVM, en cada celda se muestran 3 filas: en la primera se muestran los valores obtenidos cuando se usa *bootstrap* pareado considerando ΔMAE como estadístico; en la segunda, el parámetro es la ΔIC ; en la última fila, el parámetro es la ΔIC_{sup} .

En la primera columna de la Tabla 5.6 se muestran los resultados obtenidos de contrastar las prestaciones de k-NN y GRNN. Si se considera ΔMAE como estadístico, existen diferencias significativas en los Modelos 1, 2, 9 y 13; el error obtenido es menor cuando se estima el número de unidades vendidas usando k-NN como modelo de respuesta promocional. Para los ejemplos de cerveza de la marca del distribuidor (Modelo 9), el límite superior del IC es significativamente menor cuando se estima utilizando k-NN.

Si se contrasta las prestaciones de los métodos k-NN y ν -SVM, en términos de MAE, es preferible considerar el método ν -SVM para estimar las unidades vendidas de los productos

Cruzcampo 33, Estrella 25 por 6 y Estrella 25 por 12. Sin embargo, la dispersión de los errores es menor cuando se diseña con k-NN. En concreto, para los Modelos 2, 8, 13 y 14, se cumple tanto para ΔIC como para ΔIC_{sup} . Para el Modelo 9, esta diferencia sólo ocurre cuando se considera como estadístico ΔIC_{sup} .

Por último, para los resultados mostrados en la tercera columna de la Tabla 5.6 y correspondientes a los valores obtenidos de realizar un test *bootstrap* entre GRNN y ν -SVM, se puede decir que con ν -SVM el MAE obtenido es significativamente menor que el obtenido con GRNN para los Modelos 1, 2, 4, 5, 9, 11 y 13. En términos de dispersión, ΔIC solo es significativamente diferente para el Modelo 11; y para el Modelo 14, el límite superior del IC es menor cuando se estima con GRNN.

5.4.3. Conclusiones

De este apartado se pueden extraer tres conclusiones diferentes. En primer lugar, el uso de un test pareados *bootstrap* para realizar la comparativa entre dos métodos varía en función del parámetro que se esté analizando: ΔMAE , ΔIC y ΔIC_{sup} . Esto se debe a que el conjunto de estas medidas de prestación da una descripción estadística más detallada de los mismos, siendo a su vez la mejor considerable no solo en términos del promedio de la figura de mérito, sino también en términos de su dispersión.

En segundo lugar, de los valores de las Tablas 5.5 y 5.6 se puede concluir que para los datos de la categoría de leche, los métodos que presentan mejor comportamiento son ν -SVM y GRNN, mientras que para los ejemplos de la base de datos de la categoría de cerveza, los métodos que mejor ajustan la salida estimada son ν -SVM y k-NN.

En tercer lugar, se comparan los resultados de los tests estadísticos con los obtenidos en el Experimento 1. Para la categoría de leche se comparan los valores de las Tablas 5.3 y 5.5. Para el Modelo 1 se observa que el mínimo MAE se obtiene para el esquema MLP, sin embargo, y debido a que en el Experimento 4 se presenta algunas mejoras sobre el mismo, se realiza su análisis en la Sección 5.6. Para el Modelo 2, se comprueba que tanto los resultados obtenidos de forma empírica como los obtenidos de realizar un remuestreo *bootstrap* demuestran que las prestaciones de ν -SVM son mejores que las de k-NN y que GRNN. Además, para este modelo, si se compara de forma pareada k-NN y GRNN se obtiene que en términos de MAE es preferible estimar con GRNN. Para el Modelo 3, los resultados obtenidos son similares a los del Modelo 2. Se añade que ν -SVM presenta menor anchura que k-NN, y que GRNN es preferible a k-NN en términos de dispersión (ΔIC y ΔIC_{sup}) siendo ambos indiferentes en promedio. Para estos 2 modelos, correspondientes a los productos de la marca Ato y de la marca del distribuidor, k-NN no ajusta bien la salida estimada a la salida real, y esto puede ser debido a que el número de promociones es menor que para el resto de productos. Para el Modelo 4, los resultados obtenidos

	k-NN vs GRNN	k-NN vs ν -SVM	GRNN vs SVM
Modelo 1	-4.07 [-6.88,-1.01] -2.42 [-8.79,12.76] -4.52 [-10.05,13.87]	0.85 [-0.61,2.43] -6.73 [-13.01,7.88] -8.02 [-13.13,6.97]	4.89 [2.08,7.85] -6.53 [-10.03,5.16] -6.89 [-7.75,6.99]
Modelo 2	-13.51 [-24.07,-4.29] -11.33 [-51.85,19.99] -11.58 [-54.64,19.94]	-0.75 [-4.40,2.89] -13.20 [-24.47,-1.24] -15.83 [-26.16,16.99]	13.01 [3.43,23.07] 0.68 [-26.16,38.10] -1.12 [-26.24,34.05]
Modelo 3	0.27 [-3.22,4.04] -12.19 [-20.08,19.58] -12.26 [-19.87,19.39]	1.00 [-2.06,4.11] -12.04 [-26.49,17.23] -14.16 [-26.16,16.99]	0.87 [-1.43,3.27] -1.61 [-17.07,10.44] -1.26 [-24.56,9.99]
Modelo 4	-1.78 [-7.06,3.04] 8.76 [-17.53,20.85] 7.09 [-24.46,16.90]	4.30 [0.96,7.34] 0.91 [-23.05,12.03] 2.28 [-22.32,14.27]	6.08 [0.92,11.16] -7.07 [-27.73,16.90] -4.54 [-28.23,19.96]
Modelo 5	-1.16 [-4.43,1.71] 7.02 [-1.55,17.34] 7.15 [-1.40,17.9]	1.88 [0.12,3.75] 0.68 [-10.08,14.4] 0.75 [-9.24,13.99]	3.36 [0.55,5.99] -6.78 [-13.52,1.50] -7.26 [-13.62,0.19]
Modelo 6	-0.93 [-9.70,7.70] -9.67 [-44.34,20.41] -9.10 [-41.06,35.71]	4.26 [1.82,6.93] -3.32 [-41.07,24.58] -1.11 [-38.83,25.56]	5.50 [-4.14,14.83] 3.35 [-33.54,31.77] 2.22 [-34.67,30.54]
Modelo 7	-18.25 [-43.59,3.21] -76.33 [-171.02,31.78] -65.33 [-171.70,30.86]	-10.67 [-35.21,11.45] -47.81 [-261.81, 63.56] -39.74 [-267.33,59.68]	7.94 [-5.48,20.40] 7.89 [-127.94,118.44] 4.36 [-131.44,106.97]
Modelo 8	0.83 [-6.60,8.17] -13.11 [-46.22,13.68] -14.18 [-45.78,12.66]	2.25 [-2.97,6.98] -19.78 [-44.14,-3.42] -18.82 [-43.37,-3.49]	1.61 [-3.86,7.24] -8.97 [-32.16,16.40] -9.29 [-32.85,12.50]
Modelo 9	-12.14 [-19.99,-4.32] -23.08 [-42.88,0.98] -26.72 [-44.40,-0.43]	1.21 [-1.28,3.70] -1.11 [-15.95,6.42] -4.16 [-43.36,-3.49]	13.30 [6.45,20.67] 17.54 [-4.85,41.78] 17.74[-5.80,42.61]
Modelo 10	-6.19 [-13.24,0.53] -11.69 [-46.93,17.99] -11.78 [-47.36, 17.90]	5.10 [0.69,10.34] -7.95 [-15.95,6.43] -7.72 [-40.56,26.49]	11.17 [5.49,17.63] 5.98 [-20.11,37.46] 6.80 [-19.34,38.27]
Modelo 11	-3.65 [-8.58,0.27] -14.88 [-43.21,1.57] -16.20 [-43.02,1.30]	2.19 [-0.73,5.30] 0.66 [-13.85,13.62] 0.08 [-15.62,12.56]	6.03 [2.68,9.41] 16.23 [1.44,33.81] 14.26 [0.17,32.23]
Modelo 12	6.47 [-20.01,30.79] 82.44 [-153.18,146.97] 84.87 [-150.24,141.64]	13.04 [-18.12,44.29] 15.96 [-174.49,109.52] 10.93 [-181.41,111.25]	5.73 [-1.99,14.50] -31.25 [-126.57,16.92] -31.83 [-132.48,7.83]
Modelo 13	-6.31 [-11.29,-1.8] -18.75 [-36.99,9.57] -18.43 [-36.83,10.45]	1.62 [-2.11,5.07] -29.97 [-44.63,-3.48] -32.83 [-48.17,-4.82]	7.90 [2.40,13.23] -11.61 [-36.42,18.84] -14.87 [-38.05,14.82]
Modelo 14	-0.71 [-8.51,6.23] 0.24 [-30.71, 14.10] 3.03 [-26.46,16.20]	3.67 [-2.24,9.36] -20.96 [-81.49,-4.77] -24.50 [-85.38,-9.32]	4.83 [-1.04,10.46] -17.78 [-51.68,5.46] -25.25 [-60.78,-1.90]

Tabla 5.6: *Bootstrap* pareado para los 14 productos de cerveza considerando 3 estadísticos diferentes: ΔMAE , ΔIC y ΔIC_{sup} . En negrita se resaltan los métodos que presentan diferencias estadísticamente significativas.

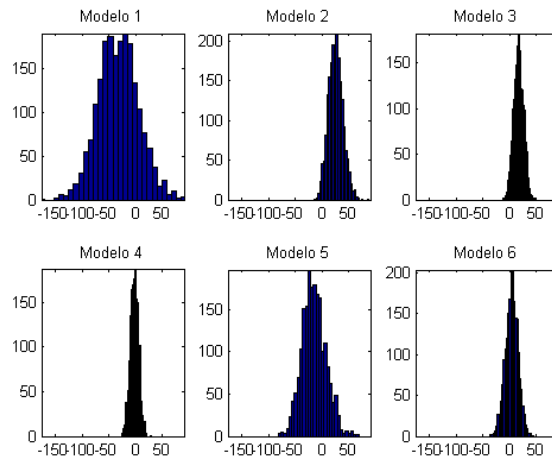
en términos absolutos se verifican cuando se realiza un remuestreo *bootstrap*. En la Tabla 5.5 se observa que ν -SVM tiene mejores prestaciones que k-NN y GRNN en términos de centrado. En la Tabla 5.3, se observa que de forma empírica k-NN es el método que presenta menor MAE. Sin embargo, si se contrasta este resultado a través de un test no paramétrico solo existen diferencias significativas con ν -SVM, y además sólo en términos de dispersión, ya que el límite superior del *IC* obtenido con el esquema k-NN es menor que el obtenido con ν -SVM. Para el Modelo 6, el mínimo valor de MAE se obtiene cuando se estima el número de unidades vendidas con GRNN, sin embargo, si se realiza un test pareado *bootstrap* no existen diferencias significativas entre ninguno de los métodos contrastados (k-NN vs GRNN, k-NN vs ν -SVM y GRNN vs ν -SVM).

Para comparar los resultados obtenidos en el Experimento 1 con los resultados de este experimento para los modelos de la categoría de cerveza es necesario analizar los valores mostrados en las Tablas 5.4 y 5.6. Del análisis empírico se había concluido que el mínimo MAE se obtenía estimando el número de unidades vendidas con ν -SVM, salvo para los Modelos 2 y 7 que k-NN presentaba mejores prestaciones. Si se contrastan los resultados a través de un remuestro *bootstrap*, los resultados obtenidos pueden ser diferentes. Así, si se compara k-NN vs GRNN, se verifican los resultados obtenidos de forma empírica. Sin embargo, para los Modelos 8 y 13, al realizar un remuestreo *bootstrap*, se obtienen que k-NN presenta mejores prestaciones que ν -SVM, lo cual difiere del análisis empírico. Lo mismo ocurre para el Modelo 14 cuando se comparan las prestaciones de GRNN y ν -SVM.

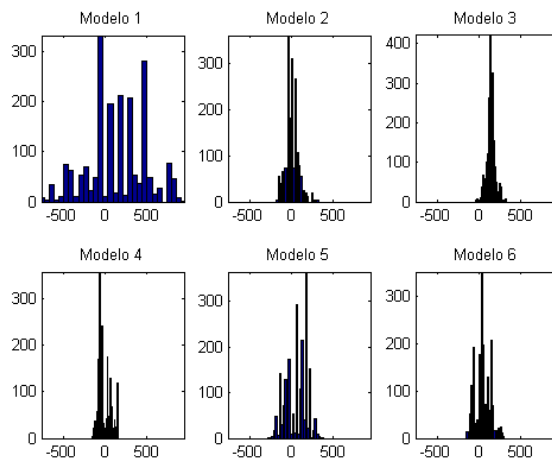
5.5. Experimento 3: ν -SVM semiparamétrica

En este apartado se realiza un análisis comparativo de la ν -SVM con núcleo RBF frente a la ν -SVM semiparamétrica, para contrastar si existen diferencias significativas entre ambas. Como ya se ha comentado, sus funciones núcleo son diferentes, ya que en la ν -SVM semiparamétrica el núcleo resultante es la suma de un núcleo lineal más un núcleo RBF (ver Sección 3.1 y Subsección 3.2.5).

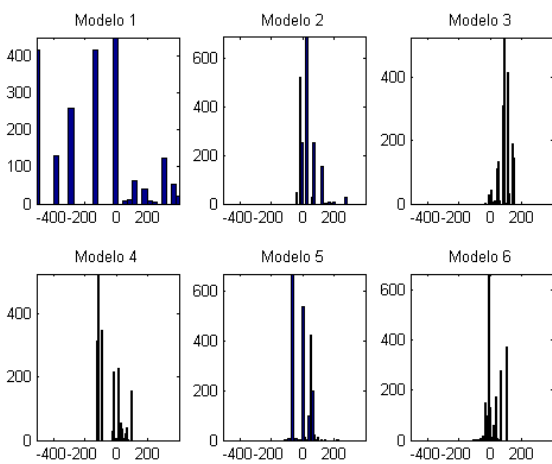
Para la categoría de leche, en cada celda de la Tabla 5.7 se muestra en la primera fila el valor de MAE empírico, MAE promedio y en la segunda fila el *IC* del 95 %. En la primera y en la segunda columna se presentan los valores obtenidos cuando se estima el número de unidades vendidas con ν -SVM y con ν -SVM semiparamétrica. En la tercera columna se muestran los resultados obtenidos cuando se realiza un test *bootstrap* para contrastar las prestaciones de ambos métodos. De las dos primeras columnas no se pueden extraer conclusiones sobre qué método ajusta mejor las estimaciones realizadas, ya que para los Modelos 2, 3 y 4, el MAE obtenido es menor cuando se diseña con ν -SVM, mientras que para el resto de modelos este error es menor cuando se modela con ν -SVM semiparamétrica. A la vista de los resultados obtenidos en las dos primeras



(a)

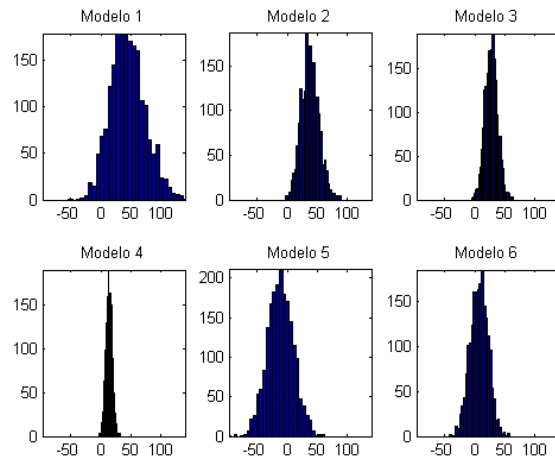


(b)

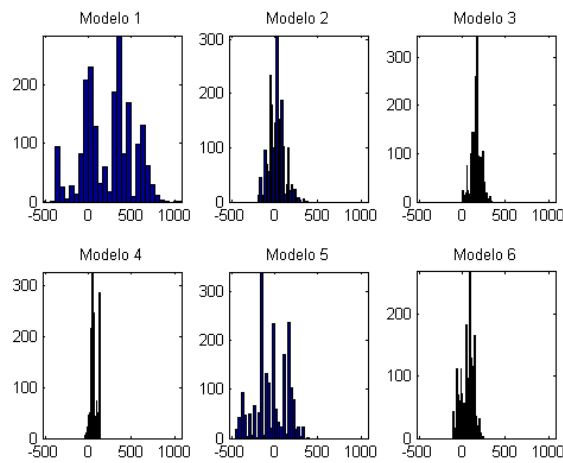


(c)

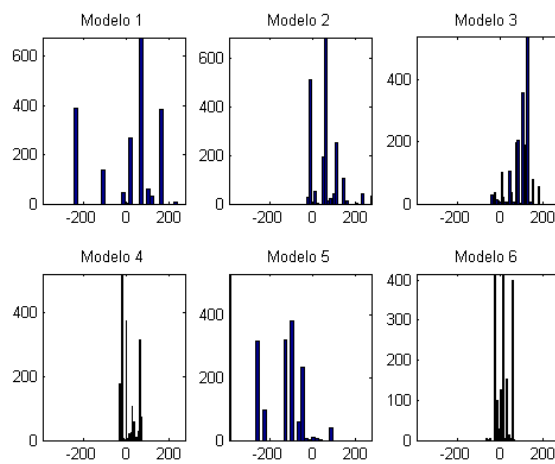
Figura 5.6: Histogramas de bootstrap pareado, k -NN vs GRNN, de diferencias en prestaciones para la categoría de leche considerando como estadístico: (a) ΔMAE ; (b) ΔIC ; (c) ΔIC_{sup} .



(a)

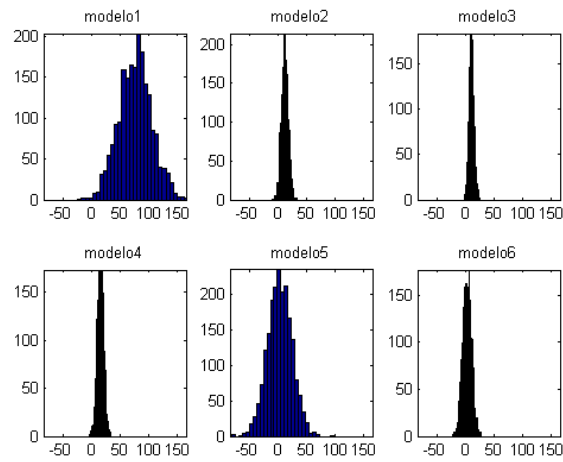


(b)

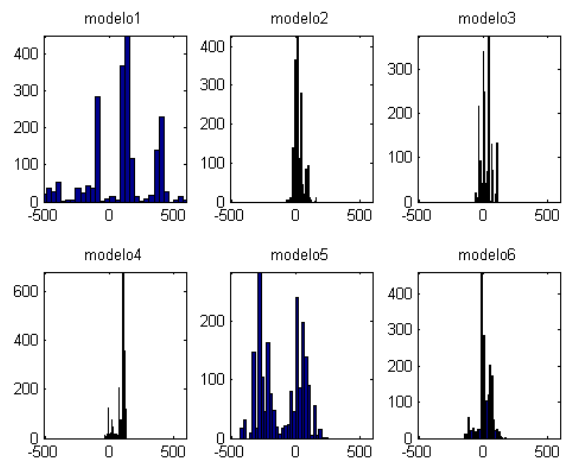


(c)

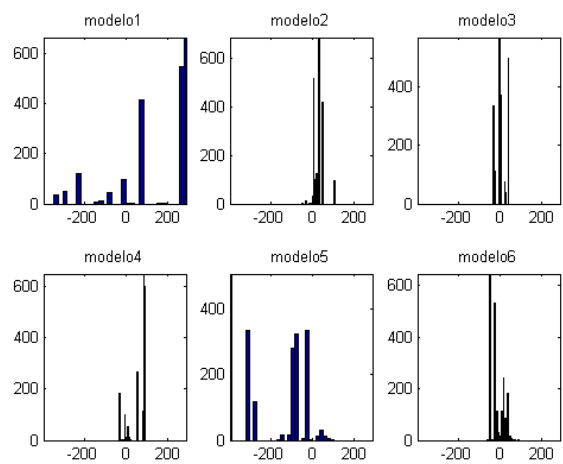
Figura 5.7: Histogramas de bootstrap pareado, k -NN vs ν -SVM, de diferencias en prestaciones para la categoría de leche considerando como estadístico: (a) ΔMAE ; (b) ΔIC ; (c) ΔIC_{sup} .



(a)



(b)



(c)

Figura 5.8: Histogramas de bootstrap pareado, GRNN vs ν -SVM, de diferencias en prestaciones para la categoría de leche considerando como estadístico: (a) ΔMAE ; (b) ΔIC ; (c) ΔIC_{sup} .

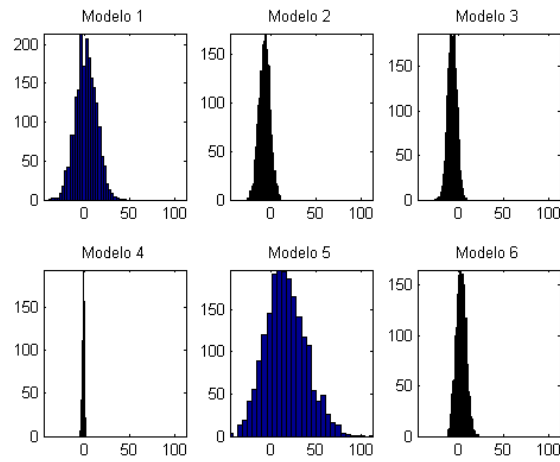
columnas de la Tabla 5.7 no se puede decir qué método es preferible en términos absolutos, con la información extraída de la tercera columna de la Tabla 5.7 y en las Figuras 5.9(a), 5.9(b) y 5.9(c), se puede concluir que los resultados obtenidos con ν -SVM semiparamétrica no son estadísticamente mejores que los obtenidos con ν -SVM con núcleo RBF.

A continuación, se muestra la comparación realizada entre ν -SVM y ν -SVM semiparamétrica para datos de la categoría de cerveza. En las dos primeras columnas de la Tabla 5.8 se recoge el valor empírico, el valor promedio y el intervalo de confianza del 95% de realizar un test no paramétrico. En términos absolutos, salvo para los Modelos 5, 7 y 8, el MAE obtenido es menor cuando no se incluye el núcleo no paramétrico. Para comprobar si existen diferencias significativas se usa un test *bootstrap*, cuyos resultados se muestran en la tercera columna de la Tabla 5.8. Al igual que ocurría para los datos de la categoría de leche, se puede decir que no existen diferencias significativas al estimar el número de unidades vendidas usando como método ν -SVM o ν -SVM semiparamétrica.

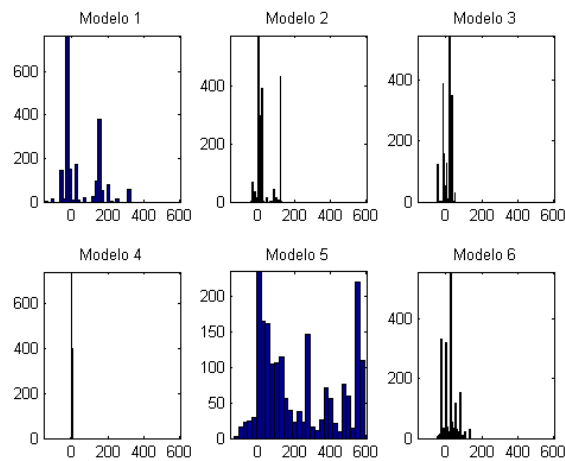
En conclusión, considerar un núcleo no paramétrico para modelizar la variable dicotómica no presenta mejoras significativas para el modelo de datos considerado en este proyecto.

	ν -SVM	ν -SVM semiparamétrica	ν -SVM vs ν -SVM semiparamétrica
Modelo 1	322.15 321.39 [230.24,420.91]	320.75 319.67 [233.59,413.37]	1.10 [-22.40,22.92] 47.65 [-57.94,326.72] 53.43 [-20.56,193.32]
Modelo 2	166.09 166.93 [117.81,218.54]	173.14 173.34 [131.49,216.88]	-7.09 [-18.64,4.75] 36.06 [-36.22,124.82] 22.80 [-15.04,108.28]
Modelo 3	136.12 135.92 [108.47,164.17]	143.16 142.46 [118.68,169.52]	-7.00 [-17.15,2.74] 5.56 [-48.77,37.73] 5.04 [-24.74,26.36]
Modelo 4	50.65 50.72 [37.47,65.45]	52.15 52.22 [39.15,67.27]	-1.50 [-3.62,0.23] -0.71 [-4.28,4.50] 0.72 [-2.27,4.50]
Modelo 5	192.60 193.64 [140.38,257.55]	175.85 175.53 [139.30,212.87]	17.65 [-23.26,66.50] 234.54 [-32.23,571.05] 272.82 [38.01,597.30]
Modelo 6	104.94 104.95 [79.99,130.59]	102.50 102.62 [78.81,125.94]	2.44 [-7.50,13.59] 17.49 [-25.86,81.98] 6.03 [-21.50,32.03]

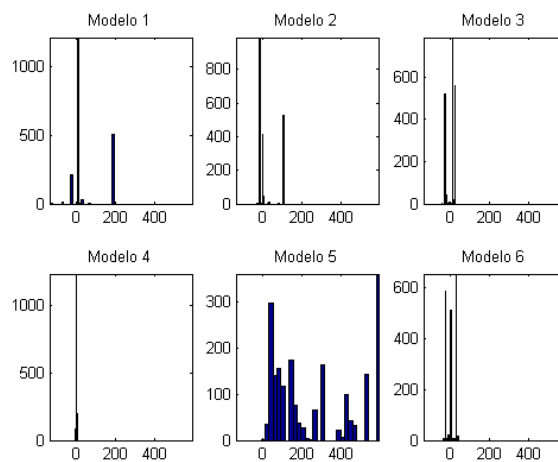
Tabla 5.7: ν -SVM vs ν -SVM semiparamétrica para leche con MAE empírico y remuestreo *bootstrap* (1^a y 2^a columna, en negrita se remarca el menor MAE); y *bootstrap* pareado con Δ MAE, Δ IC y Δ IC_{sup} (3^a columna, en negrita se remarcan las diferencias significativas).



(a)



(b)



(c)

Figura 5.9: Histogramas de bootstrap pareado de ν -SVM y ν -SVM semiparamétrica para la categoría de leche considerando como estadísticos: (a) ΔMAE ; (b) ΔIC ; (c) ΔIC_{sup} .

	ν -SVM	ν -SVM semiparamétrica	ν -SVM vs ν -SVM-semi
Modelo 1	6.98 7.03 [5.02,9.18]	7.00 7.00 [4.86,9.17]	-0.02 [-0.14,0.09] 0.23 [-0.25,0.98] 0.11 [-0.00,0.39]
Modelo 2	34.53 34.60 [26.98,42.54]	34.53 34.51 [27.37,42.32]	0.01 [-0.01,0.03] 0.00 [0.00,0.00] 0.00 [0.00,0.00]
Modelo 3	14.09 14.09 [10.46,18.48]	14.23 14.31 [10.78,18.33]	-0.14 [-0.49,0.20] 1.23 [-0.91,4.55] 0.01 [-1.27,0.92]
Modelo 4	21.41 21.22 [15.94,26.70]	21.88 21.86 [17.23,27.37]	-0.48 [-1.69,0.74] 3.46 [-5.95,6.01] 1.84 [-3.05,3.00]
Modelo 5	10.28 10.28 [7.42,13.22]	10.10 10.11 [7.58,12.94]	0.17 [-0.17,0.55] -0.01 [-1.29,3.37] 0.06 [-1.29,3.37]
Modelo 6	26.83 27.10 [19.17,35.64]	28.00 27.96 [20.14,36.16]	-1.21 [-3.24,0.00] -0.46 [-6.53,0.09] -0.47 [-6.53,0.00]
Modelo 7	89.51 88.80 [65.22,117.54]	82.57 82.62 [61.04,106.54]	6.83 [-5.71,20.03] 33.81 [-45.25,155.85] 24.08 [-48.25,127.39]
Modelo 8	34.59 34.50 [26.96,42.45]	34.39 34.36 [26.52,42.07]	0.21 [-0.29,0.72] 0.37 [-2.97,1.52] 0.40 [-2.97,1.52]
Modelo 9	19.58 19.54 [14.69,25.13]	20.60 20.69 [15.65,26.67]	-1.00 [-3.16,1.00] -1.30 [-18.56,11.55] -1.62 [-10.48,6.43]
Modelo 10	12.85 12.81 [8.11,19.11]	12.72 12.83 [8.13,19.13]	0.12 [-0.28,0.52] 0.37 [-3.78,3.31] 0.80 [-2.01,3.31]
Modelo 11	11.49 11.47 [8.20,15.26]	12.18 12.13 [9.08,15.55]	-0.67 [-2.94,1.88] -2.79 [-20.38,17.06] 8.02 [-5.70,21.13]
Modelo 12	96.85 96.29 [69.64,124.97]	115.08 115.92 [92.61,142.99]	-18.06 [-35.09,0.13] 12.27 [-30.10,31.56] 57.43 [29.65,70.20]
Modelo 13	14.92 14.93 [11.26,18.50]	15.31 15.28 [12.47,18.34]	-0.39 [-1.89,0.96] 11.09 [-2.06,17.83] 0.51 [-1.57,2.19]
Modelo 14	33.17 33.16 [26.05,40.32]	36.38 36.54 [29.17,44.58]	-3.19 [-8.09,1.72] -0.13 [-32.67,23.50] -0.84 [-16.50,16.17]

Tabla 5.8: ν -SVM vs ν -SVM semiparamétrica para leche con MAE empírico y remuestreo bootstrap (1^a y 2^a columna, en negrita se remarca el menor MAE); y bootstrap pareado con Δ MAE, Δ IC y Δ IC_{sup} (3^a columna, en negrita se remarcan las diferencias significativas).

5.6. Experimento 4: MLP, salida múltiple y salidas individuales

En este apartado se analizan dos formas de diseñar el MLP, con salidas individuales y con salida conjunta, lo cual permitirá concluir si la consideración de la información conjunta en la salida proporciona ventajas en prestaciones frente a modelos basados en ANN con salidas desacopladas. Además, se analizan las prestaciones obtenidas con diferentes funciones de activación, en concreto, la lineal y la logística sigmoideal (ver Sección 3.2.2). En primer lugar, se ha realizado la búsqueda de los parámetros libres para el nuevo modelo con salidas individuales, y se ha comparado con el modelo conjunto. En segundo lugar, se recogen en dos tablas, para leche y cerveza respectivamente, el MAE empírico, el MAE promedio obtenido con *bootstrap* y el intervalo del confianza del 95 % tanto para el modelo con salida conjunta como para el modelo con salidas individuales. Por último, se usa el test *bootstrap* pareado para contrastar si alguno de los dos métodos es significativamente mejor que el otro, en función de ΔMAE , de ΔIC y de ΔIC_{sup} .

5.6.1. Salida múltiple vs salidas individuales

El modelo de datos cuando se diseña el MLP con salidas individuales tiene 8 variables de entrada: 6 variables métricas correspondientes a los índices de precios de cada producto, dos variables dicotómicas (indicador de descuento directo e indicador de la estacionalidad); y una variable de salida (número de unidades vendidas) para cada uno de los modelos.

Parámetros libres

La búsqueda de los parámetros libres del MLP va a ser diferente cuando se tiene en cuenta salidas individuales o salida conjunta. Primero, se ha estudiado con salidas y entradas múltiples, es decir, se ha entrenado y se ha validado teniendo en cuenta los 6 modelos diferentes correspondientes a los 6 productos con los que se trabaja a lo largo del proyecto. Posteriormente, se ha estimado el número de unidades vendidas usando una red MLP diferente para cada uno de los 6 modelos. Esto ocasiona que en el caso del MLP con entradas y salidas múltiples sólo se disponga de un número óptimo de neuronas válido para todo el modelo, mientras que, cuando se trabaja con un modelo con salidas individuales, el número de neuronas en la capa oculta se calcula de forma independiente para cada modelo.

Categoría de leche.

En los datos de leche, se ha elegido aquel número de neuronas que minimiza el MAE entre la salida deseada y la salida estimada. Tanto para MLP con salida conjunta como con salidas individuales, la búsqueda del número de neuronas se ha realizado en el intervalo [1,20]. En la Figura 5.10(a) se muestra un ejemplo de búsqueda del parámetro libre (número de neuronas en la capa oculta) del MLP en los modelos de la categoría de leche. En la Tabla 5.9 se observan estos valores tanto para MLP con salida conjunta como para MLP con salidas individuales. Se

observa que el número de neuronas necesario para entrenar el MLP es mayor en el caso del MLP conjunto que el que se obtiene para cada uno de los 6 modelos de forma independiente. Cabe destacar que, en ambos casos, tanto las entradas como las salidas están normalizadas, lo cual es muy importante en el modelo con salidas individuales debido a que el volumen del número de unidades vendidas difiere considerablemente de un modelo a otro.

	n_o (MLP)	n_o (MLP_ind)
Modelo 1	17	15
Modelo 2	17	8
Modelo 3	17	1
Modelo 4	17	6
Modelo 5	17	8
Modelo 6	17	14

Tabla 5.9: Ejemplo de búsqueda realizada para los parámetros libres n_o para MLP con salida conjunta y con salidas individuales en los modelos de la categoría de leche.

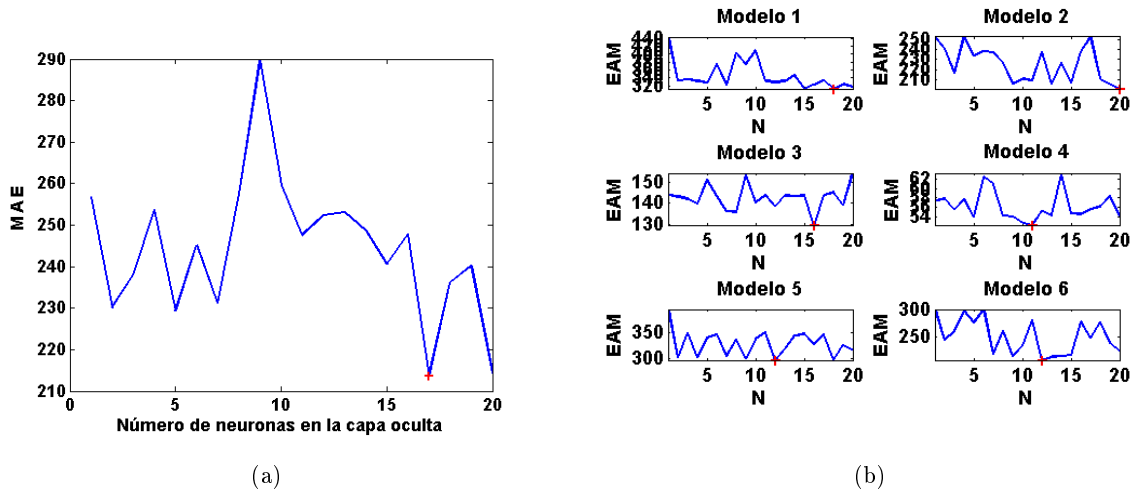


Figura 5.10: Evolución del MAE en función del número de neuronas de la capa oculta para la categoría de leche: (a) salida múltiple; (b) salidas individuales.

Categoría de cerveza.

En la Figura 5.11(a) y 5.11(b) se muestran los parámetros libres tanto para el caso de MLP con salida conjunta como con salidas individuales para los modelos de la categoría de cerveza. Estos valores se observan en la Tabla 5.10, donde se puede comprobar que el número de neuronas de la capa oculta en el caso de que se trabaje con un modelo conjunto es 8, mientras que para modelos individualizados este número puede ser mayor o menor dependiendo del modelo.

Se puede concluir que cuando se diseña el MLP con salida conjunta, la red obtenida es más

sencilla (el número de neuronas de la capa oculta es menor) para los datos de la categoría de cerveza, a pesar de analizar más productos. Si se diseña el MLP con salidas inividuales, el número de neuronas de la capa oculta depende de las características propias de cada producto.

	n_o (MLP)	n_o (MLP_ind)
Modelo 1	8	12
Modelo 2	8	8
Modelo 3	8	2
Modelo 4	8	19
Modelo 5	8	9
Modelo 6	8	8
Modelo 7	8	20
Modelo 8	8	19
Modelo 9	8	8
Modelo 10	8	20
Modelo 11	8	20
Modelo 12	8	2
Modelo 13	8	20
Modelo 14	8	5

Tabla 5.10: Ejemplo de búsqueda realizada para los parámetros libres n_o para el MLP con salidas conjunta y con salidas inividuales en función del MAE en los modelos de la categoría de cerveza.

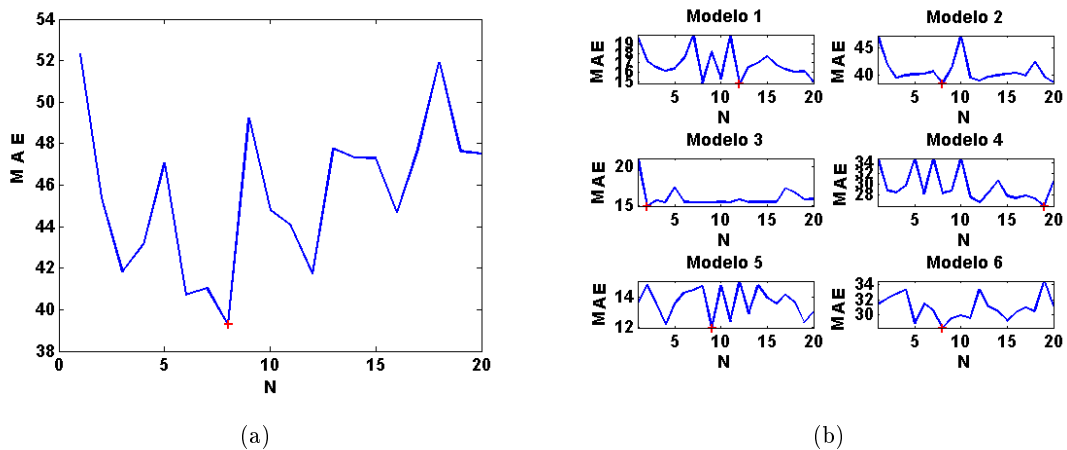


Figura 5.11: Evolución del MAE en función del número de neuronas de la capa oculta para la categoría de cerveza: (a) salida múltiple; (b) salidas individuales (para los 6 primeros modelos).

Comparativa en términos de MAE con test *bootstrap*

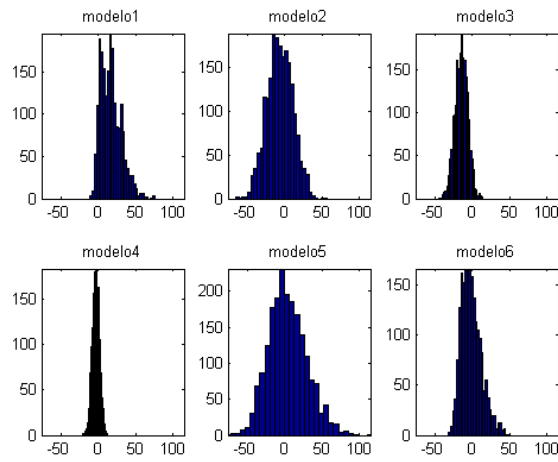
Para comparar las estimaciones del número de unidades vendidas cuando se modeliza el MLP con salidas individuales y con salida conjunta, se calcula el MAE empírico, el MAE promedio con *bootstrap* y el intervalo de confianza del 95 %.

En las dos primeras columnas de la Tabla 5.11 se muestran las estimaciones realizadas como resultado de modelizar la promoción en leche, tanto con salidas inividuales como conjunta, en términos de MAE empírico, MAE promedio e intervalo de confianza usando un test *bootstrap*. Se observa que, para los Modelos 3 y 4, las prestaciones obtenidas considerando salidas individuales son mejores que cuando se considera salida conjunta. Sin embargo, para los Modelos 1, 2, 5 y 6 es preferible incluir como entrada los índices de precios y las unidades vendidas deseadas de los diferentes modelos. No obstante, si se observa la tercera columna de la Tabla 5.11, sólo existen diferencias significativas en los Modelos 5 y 6, para los cuales tanto en promedio como es dispersión es preferible estimar el número de unidades vendidas considerando salida conjunta.

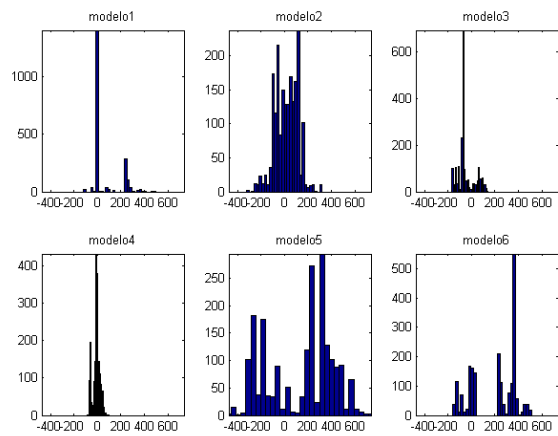
Para los productos de la categoría de cerveza, en la Tabla 5.12 se muestra en términos de MAE empírico, MAE promedio e intervalo de confianza usando un test *bootstrap* cuando se estima considerando MLP con salidas individuales (primera columna) y MLP con salida conjunta (segunda columna). De los 14 productos, sólo para los Modelos 2 y 14 se obtiene que el error empírico obtenido con MLP con salida conjunta es menor que el obtenido con salidas individuales, sin embargo, estos métodos no presentan diferencias estadísticamente significativas. Para los Modelos 1, 5, 6, 7, 8, 9 y 12, de forma empírica, es preferible estimar el número de unidades vendidas considerando salidas individuales, no obstante, cuando se realiza un remuestreo *bootstrap* con un nivel del confianza del 95 %, no existen diferencias significativas entre ambos métodos. Por otra parte, la estimación obtenida para los Modelos 3, 4, 11 y 13, analizada ésta tanto empírica como estadísticamente, mejora tanto en términos de promedio como de dispersión cuando se utiliza MLP con salidas individuales. El límite superior del intervalo de confianza para los errores de estimar el Modelo 10 es menor cuando se consideran salidas individuales.

En conclusión, para los datos de los productos promocionados de la categoría de leche es preferible considerar el MLP con salida conjunta, al menos para aquellos productos que presentan un comportamiento más estable (véase Sección 4.4). Para los productos de cerveza, en los cuales el número de unidades vendidas es considerablemente menor que para los productos promocionados de leche, MLP con salidas individuales presenta mejores prestaciones que con salida conjunta. Por lo tanto, los resultados obtenidos dependen de la categoría de producto analizada.

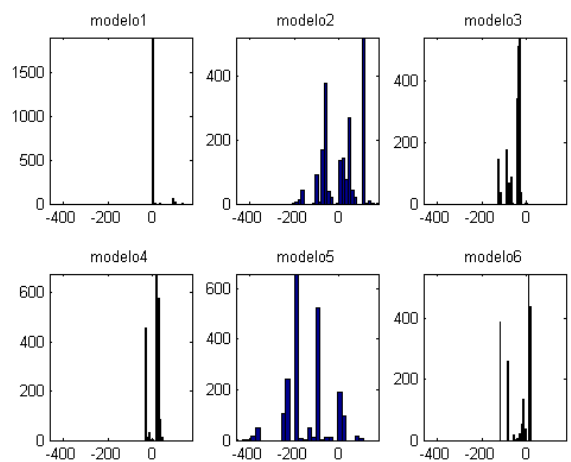
En la Figura 5.12(a), 5.12(b) y 5.12(c) se muestra un ejemplo de los histogramas obtenidos al realizar un test pareado entre MLP con salidas individuales y salida conjunta con tres estadísticos para la base de datos de leche.



(a)



(b)



(c)

Figura 5.12: Histogramas de bootstrap pareado de MLP con salida conjunta y con salidas individuales la categoría de leche con los estadísticos: (a) ΔMAE , (b) ΔIC , (c) ΔIC_{sup} .

	MLP_ind	MLP	MLP_ind vs MLP
Modelo 1	357.38 359.08 [266.27,463.26]	320.69 321.62 [248.59,401.57]	66.47 [-193.92,577.18] 37.00 [-50.81,130.32] 65.41 [-192.09,569.97]
Modelo 2	222.68 222.30 [180.80,267.25]	199.66 200.48 [157.10,247.94]	26.21 [-114.16,192.08] 28.63 [-17.91,72.95] 50.27 [-69.44,217.67]
Modelo 3	135.81 136.29 [105.50,167.31]	152.52 152.48 [119.19,188.66]	-66.91 [-144.67,49.33] -14.45 [-36.01,7.83] -57.59 [-124.63,62.07]
Modelo 4	59.62 59.76 [44.25,77.40]	72.29 72.4 [58.51,88.34]	-7.79 [-49.20,18.67] -12.79 [-27.25,1.37] -13.44 [-54.50,7.05]
Modelo 5	305.20 310.86 [227.98,397.54]	198.71 197.83 [148.57,250.74]	258.84 [63.98,528.11] 103.33 [21.01,186.19] 267.33 [74.95,544.11]
Modelo 6	226.69 226.70 [167.35,293.25]	125.62 125.04 [97.81,155.32]	304.52 [98.22,447.10] 100.70 [37.84,163.13] 304.50 [108.68,473.25]

Tabla 5.11: *MLP salida conjunta vs salidas individuales para leche con MAE empírico y remuestreo bootstrap (1^a y 2^a columna, en negrita se remarca el menor MAE); y bootstrap pareado con: ΔMAE , ΔIC y ΔIC_{sup} (3^a columna, en negrita se remarcen las diferencias significativas).*

Comparativa con el resto de métodos

En el Experimento 2 se ha realizado la comparación entre los métodos k-NN vs GRNN, k-NN vs ν -SVM y GRNN vs ν -SVM. Para completar dicho análisis, en este apartado se comprueba de forma pareada qué método presenta mejores prestaciones. En concreto se contrasta el MAE obtenido tras modelizar el MLP con salidas individuales con k-NN, GRNN, y con ν -SVM. Como en el Experimento 2, se realiza *bootstrap* con tres figura de mérito distintas: ΔMAE , ΔIC Y ΔIC_{sup} .

En la Tabla 5.13 se muestran la diferencias existentes en la base de datos de leche entre MLP con salidas individuales frente a k-NN, GRNN y ν -SVM. Se observa que en términos de MAE, solo existen diferencias significativas para el Modelo 4, siendo menor el error obtenido con ν -SVM. En términos de dispersión, existen diferencias significativas en los Modelos 4 y 6. Para el Modelo 4 la dispersión obtenida con ν -SVM es menor que la obtenida con MLP con salidas individuales; para el Modelo 6, en términos de dispersión es preferible crear el modelo promocional con GRNN que con MLP con salidas individuales.

Los resultados de la Tabla 5.14 verifican los resultados obtenidos en el Experimento 2 para los datos de la categoría de cerveza, los métodos que mejor ajustan la salida obtenida son ν -SVM

	MLP_ind	MLP	MLP_ind vs MLP
Modelo 1	16.59 16.63 [13.29,20.35]	18.12 18.17 [14.75,21.62]	-1.58 [-4.33,0.99] -7.74 [-23.71,17.30] -7.55 [-23.71,17.30]
Modelo 2	45.50 45.51 [35.64,55.84]	44.93 45.24 [36.31,53.60]	0.71 [-5.74,7.53] 9.10 [-14.20,47.18] 8.46 [-14.20,44.95]
Modelo 3	16.08 16.12 [12.79,20.32]	18.643 18.32 [13.80,24.11]	-2.56 [-5.74,-0.10] -17.25 [-29.84,-1.50] -15.57 [-28.16,-0.00]
Modelo 4	29.52 29.48 [23.95,35.75]	39.55 38.72 [32.13,45.89]	-10.16 [-15.41,-4.77] -20.67 [-43.83,-1.10] -21.04 [-43.83,-2.55]
Modelo 5	14.76 14.73 [11.66,17.85]	15.22 15.37 [11.78,19.36]	-0.51 [-2.89,1.83] -7.24 [-13.49,-0.00] -7.36 [-13.40,-0.00]
Modelo 6	31.83 31.96 [25.67,38.13]	36.79 38.18 [30.35,46.95]	-5.00 [-10.40,-0.07] -19.05 [-44.74,13.75] -20.08 [-44.55,11.16]
Modelo 7	125.58 125.29 [102.40,148.82]	134.75 137.96 [107.54,170.35]	-8.50 [-28.70,9.85] -88.04 [-187.24,3.00] -83.85 [-187.23,3.00]
Modelo 8	36.87 37.05 [28.23,45.81]	40.88 40.11 [30.01,50.75]	-3.92 [-10.49,1.77] -12.29 [-41.54,2.56] -13.15 [-44.29,2.56]
Modelo 9	30.45 30.60 [25.09,36.11]	33.69 3.94 [27.93,39.85]	-3.16 [-9.67,2.68] -8.74 [-37.67,16.12] -10.27 [-38.40,15.38]
Modelo 10	34.31 34.26 [26.78,43.03]	37.85 37.84 [27.75,48.53]	-3.37 [-8.80,2.07] -19.42 [-47.26,2.69] -19.94 [-47.26,-0.00]
Modelo 11	17.68 17.68 [13.57,22.17]	22.32 21.72 [15.98,28.93]	-4.63 [-8.34,-1.54] -25.58 [-31.12,-6.50] -25.14 [-31.12,-6.37]
Modelo 12	101.95 102.14 [77.47,127.91]	109.09 108.08 [80.90,139.97]	-7.31 [-25.66,9.12] -28.14 [-133.09,52.95] -32.22 [-133.09,52.95]
Modelo 13	33.55 33.61 [28.26,39.66]	41.40 41.00 [33.53,49.26]	-7.85 [-14.36,-1.48] -39.14 [-53.13,-6.91] -38.28 [-51.15,-6.91]
Modelo 14	50.23 50.32 [40.64,60.81]	48.26 47.67 [38.09,58.75]	1.85 [-4.23,8.26] -14.24 [-49.84,17.17] -13.91 [-50.15,17.16]

Tabla 5.12: *MLP salida conjunta vs salidas individuales para cerveza con MAE empírico y remuestreo bootstrap (1^ª y 2^ª columna, en negrita se remarca el menor MAE); y bootstrap pareado con: ΔMAE , ΔIC y ΔIC_{sup} (3^ª columna, en negrita se remarcan las diferencias significativas).*

	k-NN vs MLP_ind	GRNN vs MLP_ind	ν -SVM vs MLP_ind
Modelo 1	28.07 [-61.06,121.80] 303.90 [-526.62,1110.87] 217.66 [-144.66,798.62]	61.89 [-29.26,152.58] 207.39 [-82.88,880.21] 169.95 [-119.20,894.82]	-12.72 [-98.70,69.68] 97.60 [-432.09,741.73] 207.95 [-45.33,684.93]
Modelo 2	12.00 [-32.83,57.67] 6.28 [-296.03,260.98] 93.75 [-173.03,304.00]	-13.99 [-54.90,24.83] 24.78 [-133.92,96.24] 27.11 [-140.22,93.90]	-26.14 [-62.63,10.95] -9.45 [-344.41,314.62] 33.64 [-235.58,317.31]
Modelo 3	21.66 [-6.81,50.17] 71.09 [-48.89,179.21] 52.50 [3.61,93.89]	4.96 [-16.93,26.80] -40.05 [-83.82,33.27] -47.45 [-87.53,39.58]	-5.80 [-28.29,16.07] -83.36 [-194.82,7.86] -42.80 [-87.10,25.64]
Modelo 4	-4.19 [-20.04,12.01] -9.81 [-113.49,97.41] -33.91 [-87.21,51.69]	-2.27 [-13.72,9.08] -2.16 [-79.80,45.17] 15.44 [-75.24,39.97]	-17.66 [-28.52,-7.14] -75.27 [-105.14,-23.73] -46.57 [-63.74,-7.71]
Modelo 5	-17.35 [-78.81,37.75] -189.69 [-597.43,334.76] 70.89 [-29.05,152.78]	-2.63 [-60.86,43.84] -89.85 [-215.29,186.17] -66.81 [-400.34,179.27]	-4.70 [-64.21,54.67] -147.80 [-574.74,495.29] 271.33 [23.19,549.02]
Modelo 6	-13.17 [-57.16,25.59] -246.46 [-518.74,100.96] -14.76 [-88.00,37.17]	-16.82 [-54.62,15.26] -58.29 [-138.41,-18.49] -58.57 [-130.05,-22.83]	-18.81 [-52.29,9.40] -300.84 [-537.63,31.78] -28.27 [-88.35,61.49]

Tabla 5.13: *Media e IC de bootstrap pareado para los productos de leche considerando ΔMAE , ΔIC y ΔIC_{sup} . En negrita se remarcan las diferencias estadísticamente significativas.*

y k-NN. En concreto, cuando se contrasta k-NN vs MLP con salidas individuales, el ajuste de los datos realizado por k-NN es significativamente mejor para los Modelos 1, 2, 7, 8, 9, 12, 13 y 14 en términos de ΔMAE ; para los Modelos 2, 9, 12, 13 y 14 en términos de ΔIC ; y para los Modelos 2, 9, 12 y 13 en términos de ΔIC_{sup} . Los resultados obtenidos de realizar un test pareado entre GRNN y MLP con salidas individuales se muestran en la segunda columna. Se observa que en términos de MAE sólo existen diferencias significativas para los Modelos 12 y 14, ajustado mejor la salida obtenido el método GRNN. También, para el Modelo 14 es preferible GRNN en términos de dispersión, mientras que para el Modelo 11 es preferible el método MLP con salidas individuales.

En la tercera columna de la Tabla 5.13 se muestran los resultados de contrastar el método ν -SVM frente al método MLP con salidas individuales. Se observa que en función del estadístico analizado es preferible un método u otro. Así, en términos de error promedio, ΔMAE , es significativamente mejor estimar con ν -SVM los Modelos 1, 2, 4, 5, 8, 9, 10, 12, 13 y 15. En términos de dispersión, en concreto, en función de ΔIC , es significativamente mejor estimar con el método MLP con salidas individuales para los Modelos 3, 5, 6, 7, 8 y 11. En función de ΔIC_{sup} es significativamente mejor la salida obtenida con ν -SVM para los Modelos 1, 9 y 14, mientras que para el Modelo 11 es preferible la salida obtenida del método MLP con salidas individuales.

	k-NN vs MLP_ind	GRNN vs MLP_ind	ν -SVM vs MLP_ind
Modelo 1	-6.37 [-10.16,-2.82] -7.65 [-22.10,7.71] -8.31 [-22.60,4.90]	-2.43 [-6.76,1.75] -8.58 [-18.57,1.57] -8.14 [-16.71,2.92]	-7.26 [-10.76,-4.04] 2.41 [-10.64,12.15] -16.91 [-30.10,-7.72]
Modelo 2	-15.75 [-27.20,-5.66] -49.93 [-78.60,-2.60] -49.97 [-79.63,-3.62]	-1.87 [-14.13,10.86] -31.61 [-63.20,20.15] -36.64 [-69.08,14.15]	-14.80 [-26.30,-4.44] 320.80 [-21.61,75.05] -58.81 [-97.29,5.88]
Modelo 3	0.73 [-1.56,2.92] 1.17 [-20.06,25.14] 1.39 [-19.87,24.53]	0.61 [-3.16,4.32] -0.14 [-20.81,21.65] -1.35 [-22.04,20.32]	-0.29 [-3.68,2.59] 29.79 [6.47,46.08] -3.85 [-26.18,13.92]
Modelo 4	-1.60 [-8.26,4.15] -11.17 [-33.20,12.97] -11.71 [-32.86,11.17]	0.04 [-6.98,6.73] -15.45 [-39.39,15.61] -15.40 [-40.46,14.41]	-6.04 [-10.86,-1.50] 17.33 [-15.36,37.81] -17.45 [-48.49,0.28]
Modelo 5	-1.29 [-3.83,1.39] 4.68 [-5.82,14.28] 4.79 [-5.33,13.38]	0.09 [-3.43,3.52] -4.64 [-16.06,6.37] -4.56 [-15.52,6.28]	-3.18 [-5.91,-0.51] 21.12 [0.23,44.88] -9.00 [-25.85,13.30]
Modelo 6	-0.68 [-7.61,5.97] 19.57 [-35.36,43.65] 20.84 [-34.25,44.76]	0.60 [-7.93,9.40] 25.32 [-21.41,55.04] 26.76 [-19.62,57.17]	-4.98 [-12.49,2.75] 97.08 [19.42,156.21] 4.27 [-47.75,41.83]
Modelo 7	-16.67 [-30.49,-2.31] -23.38 [-114.56,69.26] -28.72 [-120.56,63.26]	1.85 [-22.93,27.99] 25.35 [-41.66,112.17] 17.19 [-46.49,106.81]	-5.68 [-33.71,20.93] 219.05 [86.53,409.23] 43.29 [-90.22,236.00]
Modelo 8	-9.38 [-19.30,-0.87] -25.25 [-50.04,8.41] -25.76 [-51.93,8.34]	-10.01 [-21.19,0.85] -18.80 [-47.92,19.65] -21.49 [-45.94,18.99]	-11.59 [-21.47,-2.70] 67.60 [17.40,107.63] -28.47 [-59.88,3.60]
Modelo 9	-9.28 [-17.29,-2.06] -33.57 [-51.56,-7.07] -32.89 [-50.20,-8.41]	2.88 [-7.94,13.26] -11.07 [-32.76,12.95] -7.54 [-30.21,17.32]	-10.35 [-18.04,-3.02] 18.79 [-20.52,43.59] -43.40 [-67.47,-28.08]
Modelo 10	-7.76 [-16.54,0.29] -8.94 [-71.47,54.93] -8.83 [-71.31,26.83]	-1.54 [-11.48,8.64] -1.15 [-57.26,45.16] -5.50 [-57.09,49.59]	-12.96 [-20.58,-6.02] -13.08 [-67.06,48.17] -45.08 [-100.32,11.38]
Modelo 11	0.67 [-2.73,4.08] 4.79 [-3.42,15.29] 4.88 [-3.22,15.48]	4.60 [-0.23,9.95] 24.75 [2.49,48.60] 24.53 [1.98,48.18]	-1.40 [-4.90,2.21] 31.32 [12.25,48.03] 3.72 [-12.19,17.54]
Modelo 12	-41.33 [-74.57,-10.99] -107.39 [-222.99,-3.87] -109.96 [-222.85,-8.51]	-47.70 [-78.69,-21.19] -150.19 [-238.90,78.40] -146.60 [-242.69,48.53]	-53.39 [-86.54,-22.93] 51.64 [-102.93,288.90] -124.90 [-262.92,109.75]
Modelo 13	-8.53 [-16.75,-1.20] -45.54 [-77.06,-4.28] -45.19 [-79.52,-3.26]	-2.08 [-10.96,6.21] -18.87 [-62.10,9.77] -15.12 [-62.86,10.72]	-10.08 [-18.40,-3.04] -10.04 [-52.37,38.73] -57.12 [-89.89,-15.85]
Modelo 14	-11.73 [-20.82,-4.11] -41.83 [-73.30,-0.71] -37.31 [-69.33,3.94]	-10.47 [-20.49,-0.53] -30.94 [-79.07,-4.18] -30.00 [-66.68,-3.84]	-15.37 [-25.01,-5.89] 9.68 [-29.98,59.02] -48.14 [-89.62,-5.41]

Tabla 5.14: *Media e IC de bootstrap pareado para los productos de cerveza considerando ΔMAE , ΔIC y ΔIC_{sup} . En negrita se remarcan las diferencias estadísticamente significativas.*

En conclusión, para los modelos de la categoría de leche el MLP con salidas individuales no presenta mejores prestaciones que el resto de métodos. Sin embargo, para algunos modelos de la categoría de cerveza, las prestaciones en términos de dispersión del MLP con salidas individuales son mejores que las de GRNN o ν -SVM.

5.6.2. Función de activación

Partiendo de la teoría presentada en la Sección 3.2.2, se analiza a continuación de forma experimental dos tipos de funciones globales implementadas en las neuronas de la capa oculta. En concreto, se analiza un ejemplo de una función de tipo lineal y una función logística sigmoideal. Se muestran los resultados obtenidos con ambas funciones cuando se estima con MLP considerando salidas individuales.

Si se analiza la categoría de producto de leche, en la Tabla 5.15, de forma empírica (primera y segunda columna) se observa que las estimaciones obtenidas son mejores cuando se considera una función lineal en las neuronas de la capa oculta. Sin embargo, sólo existen diferencias significativas en los Modelos 5 y 6.

	Individual (Sigmoideal)	Individual (Lineal)	Sigmoideal vs Lineal
Modelo 1	358.26 368.36 [258.02,461.30]	325.72 326.34 [252.42,414.28]	31.56 [-66.70,138.97] 140.33 [-128.92,583.36] 141.05 [-118.15,584.03]
Modelo 2	235.09 235.50 [195.20,281.78]	203.87 205.62 [156.88,256.22]	32.00 [-4.58,67.50] -36.56 [-219.90,85.88] -25.61 [-185.48,91.34]
Modelo 3	143.31 148.67 [111.92,174.92]	137.72 140.73 [106.40,171.07]	6.12 [-14.09,29.12] 2.24 [-50.53,84.28] 12.51 [-36.64,87.05]
Modelo 4	56.92 58.43 [40.36,78.24]	64.75 66.43 [48.63,82.63]	-8.97 [-18.88,2.04] 0.79 [-28.63,56.75] 2.43 [-29.00,54.28]
Modelo 5	347.79 360.08 [271.89,426.30]	191.91 190.17 [154.73,231.23]	153.45 [68.10,242.22] 372.88 [121.15,555.56] 377.80 [116.54,549.46]
Modelo 6	245.50 255.17 [190.68,302.30]	116.35 116.12 [90.41,142.95]	127.18 [69.19,187.44] 335.52 [134.81,445.85] 346.84 [164.44,465.16]

Tabla 5.15: MAE obtenido a la salida del MLP según la función de las neuronas de la capa oculta para la categoría de leche.

Para la categoría de cerveza, en Tabla 5.16 se muestra que en términos absolutos, es mejor considerar una función lineal en las neuronas de la capa oculta para los Modelos 1, 2, 3, 4, 6, 7, 10, 11 y 13. Sin embargo, en términos relativos, resultado de realizar un test no paramétrico, existen diferencias estadísticamente significativas en los Modelos 8, 9, 12, 11 y 13. Para los Modelos 8, 9 y 12 es mejor implementar una función sigmoïdal en las neuronas de la capa oculta, mientras que para los Modelos 11 y 13 es preferible implementar una función lineal.

Se puede concluir que para la categoría de leche es mejor entrenar el MLP considerando una función lineal en las neuronas de la capa oculta. En el caso de cerveza, para productos con pocas unidades vendidas como son los Modelos 11 y 13, es mejor considerar una función lineal, mientras que para productos en los que el volumen de ventas es mayor, las prestaciones mejoran cuando se considera una función sigmoïdal.

5.7. Experimento 5: selección de variables

Teniendo en cuenta dos de los diferentes elementos considerados en el modelo de datos, a saber, la variable dicotómica y la línea de base, en este apartado se propone realizar un estudio específico de selección de variables, para contrastar con cuál de ellos se obtiene menor MAE, y se ha comprobado para cada uno de los métodos estudiados en el trabajo (k-NN, GRNN, MLP y ν -SVM). Las posibles opciones que se comprueban son:

- Caso 1: Modelo con línea de base y con variable dicotómicas (LB - VD).
- Caso 2: Modelo con línea de base y sin incluir variables dicotómicas (LB - NoVD).
- Caso 3: Modelo sin incluir línea de base pero con variables dicotómicas (NoLB - VD).
- Caso 4: Modelo sin incluir ni línea de base ni variables dicotómicas (NoLB - NoVD).

Para la categoría de leche se obtienen los siguientes resultados:

- En la Tabla 5.17 se muestra para k-NN las variaciones respecto al uso o no de variables dicotómicas. Como se observa, para los Modelos 1, 2 y 3 es mejor incluir la variable dicotómica pero no incluir la línea de base. Sin embargo, para los Modelos 5 y 6 es mejor incluir la línea de base y no incluir en el modelo la información relativa a la variable dicotómica.
- En la Tabla 5.18 se observan los resultados obtenidos si se utiliza como método GRNN. Como se observa, para los Modelos 1, 2 y 4, el mínimo MAE se obtiene cuando no se incluye ni línea de base ni variable dicotómica. Sin embargo, para los Modelos 5 y 6 los resultados obtenidos en términos de MAE son mejores cuando se incluyen como variables del modelo tanto la variable dicotómica como la línea de base.

	Individual (Sigmoidal)	Individual (Lineal)	Sigmoidal vs Lineal
Modelo 1	17.95 17.99 [14.97,21.15]	13.16 13.24 [9.70,16.94]	4.60 [-0.46,8.74] -0.28 [-17.03,11.48] 0.54 [-16.77,10.62]
Modelo 2	47.77 48.04 [37.98,57.85]	44.20 44.19 [36.89,52.84]	3.49 [-6.30,14.64] 22.72 [-9.64,69.43] 17.86 [-12.64,57.97]
Modelo 3	16.50 16.11 [12.56,21.26]	14.54 14.55 [10.71,18.94]	2.01 [-2.33,5.72] -3.92 [-20.32,13.80] -4.95 [-21.54,12.96]
Modelo 4	28.93 28.41 [23.11,35.03]	26.71 26.80 [20.01,34.16]	2.14 [-5.93,9.63] -15.54 [-38.31,12.80] -15.50 [-39.91,13.69]
Modelo 5	13.00 13.31 [10.04,16.37]	14.68 14.74 [11.25,18.50]	-1.72 [-5.40,1.90] -7.95 [-21.41,4.82] -7.48 [-20.86,4.45]
Modelo 6	32.63 34.37 [27.29,37.69]	29.34 29.46 [23.44,35.67]	3.29 [-4.66,9.80] -7.92 [-32.28,5.08] -3.97 [-30.05,8.75]
Modelo 7	118.06 116.81 [96.80,141.15]	94.82 94.90 [70.09,119.85]	22.97 [-9.20,52.94] -23.62 [-80.67,74.21] -18.75 [-78.52,77.48]
Modelo 8	37.19 36.65 [29.19,46.09]	48.28 48.47 [38.25,59.22]	-11.48 [-23.66,-0.18] -25.41 [-81.01,-0.99] -25.51 [-71.11,-2.59]
Modelo 9	27.75 28.13 [22.92,32.76]	27.89 27.83 [20.15,36.46]	0.07 [-9.34,9.23] -28.13 [-42.41,-1.93] -26.39 [-39.50,-1.65]
Modelo 10	31.00 31.09 [23.00,39.86]	24.91 24.93 [16.75,34.42]	6.22 [-4.35,16.38] 8.43 [-55.92,51.01] 8.48 [-54.53,51.97]
Modelo 11	20.18 19.78 [14.73,26.71]	12.44 12.46 [9.30,15.92]	7.73 [2.92,13.74] 33.34 [-4.94,57.47] 32.47 [-14.83,57.04]
Modelo 12	109.11 108.40 [80.93,139.88]	136.16 135.87 [103.36,171.17]	-27.44 [-53.69,-1.93] -76.56 [-213.62,149.96] -74.79 [-207.25,148.53]
Modelo 13	36.96 36.42 [30.60,43.96]	24.07 24.00 [17.06,32.78]	12.57 [3.65,21.21] -8.78 [-59.77,24.44] -9.42 [-59.79,18.65]
Modelo 14	45.57 45.29 [35.41,56.31]	50.33 50.22 [35.98,67.58]	-4.41 [-18.32,7.52] -68.39 [-100.58,12.00] -71.95 [-99.80,-5.07]

Tabla 5.16: Error absoluto medio obtenido a la salida del MLP según la función de las neuronas de la capa oculta para la categoría de cerveza.

- Los resultados obtenidos sobre la inclusión de la línea de base y la variable dicotómica cuando se modeliza con MLP con salidas individuales se muestran en la Tabla 5.19. Se observa que, para los Modelos 2, 5 y 6, es mejor incluir la línea de base y no incluir la variable dicotómica.
- En la Tabla 5.20 se muestran los resultados obtenidos de incluir o no la línea de base cuando se modeliza usando ν -SVM. Como para el resto de métodos, no se puede decir si es preferible incluir la línea de base y/o la variable dicotómica, ya que depende del modelo a analizar.

Como se observa en las Tablas 5.17, 5.18, 5.19 y 5.20, correspondientes a los métodos k-NN, GRNN, MLP modelos individualizados y ν -SVM, respectivamente, no es sencillo decidir para cada método si es mejor o no el uso de la variable dicotómica y de la línea de base. Por ejemplo, para ν -SVM se puede decir que existe un mayor número de modelos en los cuales se obtiene menor MAE incluyendo como variables de entrada tanto línea de base como variables dicotómicas; para k-NN se obtiene mejores resultados cuando no se usa línea de base pero sí la variable dicotómica; cuando se usa GRNN es mejor no incluir para modelizar el modelo ni línea de base, ni variables dicotómicas; y para MLP son salidas individualizadas los mejores resultados se obtienen incluyendo como variables del modelo la línea de base, pero no incluyendo variables dicotómicas. Debido a que la diferencia existente en términos de MAE (primera línea de cada celda, primer valor) no es muy significativa, para decidir sobre la inclusión de la línea de base y/o variables dicotómicas, se emplea el remuestreo *bootstrap*. El resultado obtenido es similar al obtenido de forma empírica, no permite decidir de forma operativa cuál es la combinación óptima en términos de incluir línea de base y/o las variables dicotómicas. Por ese motivo, se decide usar un test pareado, *bootstrap*, que permita tomar esa elección de forma adecuada con un nivel de confianza del 95 %.

	LB - VD	LB - NoVD	NoLB - VD	NoLB -NoVD
Modelo 1	366.44 365.50 [269.95,465.16]	366.44 364.56 [269.49,475.58]	353.82 349.08 [250.24,453.78]	353.82 352.47 [256.74,463,29]
Modelo 2	205.21 205.56 [156.94,257.97]	199.99 199.66 [152.15,249.14]	192.36 192.08 [145.42,239.44]	194.57 195.29 [147.05,250.39]
Modelo 3	164.62 165.11 [131.26,199.64]	163.79 163.57 [129.95,199.31]	159.97 159.82 [126.73,197.65]	163.83 163.58 [129.28,200.09]
Modelo 4	63.27 63.18 [48.61,79.27]	62.74 62.72 [48.14,78.43]	61.98 62.14 [47.02,77.76]	61.98 61.38 [46.99,75.89]
Modelo 5	178.99 178.49 [131.93,226.97]	178.01 178.05 [133.70,228.13]	239.54 238.24 [174.73,307.64]	239.86 239.55 [172.04,312.32]
Modelo 6	119.68 119.54 [92.13,148.40]	111.07 111.19 [85.74,139.31]	166.60 167.09 [112.54,231.97]	171.64 172.77 [121.14,229.52]

Tabla 5.17: *k*-NN. Inclusión de variables dicotómicas y línea de base en términos de MAE empírico y remuestreo bootstrap para los diferentes métodos y para los productos de la categoría de leche. En negrita se ramarca para cada producto el método con menor MAE.

	LB - VD	LB - NoVD	NoLB - VD	NoLB -NoVD
Modelo 1	412.42 410.61 [314.41,532.60]	395.74 397.96 [292.77,521.12]	387.71 390.69 [296.09,494.89]	382.80 384.88 [286.83,500.35]
Modelo 2	187.57 187.59 [139.29,235.98]	199.21 199.69 [149.90,257.21]	185.51 185.10 [144.55,230.59]	185.16 184.95 [143.84,232.89]
Modelo 3	150.77 150.68 [125.00,177.73]	144.99 145.31 [117.97,172.55]	150.96 151.19 [124.13,179.92]	148.93 149.08 [122.72,179.37]
Modelo 4	63.89 63.53 [47.76,83.19]	63.28 63.20 [47.35,84.22]	59.86 60.27 [46.14,77.00]	59.46 59.23 [45.62,75.19]
Modelo 5	193.34 194.21 [154.14,238.14]	196.89 196.58 [151.38,246.64]	220.93 220.14 [156.71,281.99]	239.27 240.14 [171.85,323.24]
Modelo 6	104.42 104.38 [78.14,131.18]	107.56 108.03 [83.32,135.26]	138.74 138.04 [102.24,180.27]	154.97 154.74 [111.49,207.07]

Tabla 5.18: GRNN. Inclusión de variables dicotómicas y línea de base en términos de MAE empírico y remuestreo bootstrap para los diferentes métodos y para los productos de la categoría de leche. En negrita se ramarca para cada producto el método con menor MAE.

	LB - VD	LB - NoVD	NoLB - VD	NoLB -NoVD
Modelo1	320.75 320.56 [247.23,407.72]	357.26 354.40 [266.16,456.82]	374.59 371.48 [271.59,481.88]	390.88 389.42 [306.58,286.09]
Modelo 2	198.89 198.22 [154.57,244.11]	188.34 188.72 [144.60,235.99]	241.46 242.58 [192.47,306.18]	207.27 206.96 [155.62,262.41]
Modelo 3	151.23 151.83 [122.15,182.68]	150.71 150.88 [117.60,184.12]	146.63 146.44 [115.79,178.59]	153.76 153.43 [121.39,189.20]
Modelo 4	72.41 72.85 [58.06,89.09]	68.68 68.70 [53.01,86.22]	64.44 64.46 [45.99,85.22]	67.48 67.63 [52.83,84.82]
Modelo 5	196.36 196.83 [151.58,248.20]	186.22 186.15 [140.73,233.59]	287.62 287.60 [218.60,368.81]	241.68 241.66 [177.55,321.45]
Modelo 6	127.68 127.64 [99.38,157.79]	122.73 122.59 [96.64,149.99]	196.12 194.43 [150.15,246.51]	183.41 183.24 [136.42,232.25]

Tabla 5.19: *MLP*. Inclusión de variables dicotómicas y línea de base en términos de MAE empírico y remuestreo bootstrap para los diferentes métodos y para los productos de la categoría de leche. En negrita se ramarca para cada producto el método con menor MAE.

	LB - VD	LB - NoVD	NoLB - VD	NoLB -NoVD
Modelo 1	322.15 320.21 [232.78,423.67]	336.98 337.31 [249.96,434.62]	293.17 293.10 [205.86,391.32]	310.78 308.51 [221.61,408.79]
Modelo 2	166.09 164.64 [120.33,215.70]	171.44 335.25 [250.39,431.68]	178.26 292.01 [208.99,378.27]	155.29 154.55 [109.80,208.72]
Modelo 3	136.12 135.86 [109.36,164.29]	139.55 139.86 [111.45,167.15]	133.56 133.51 [104.30,164.12]	136.95 136.80 [107.47,167.15]
Modelo 4	50.65 50.40 [37.23,64.73]	52.47 52.66 [39.52,67.61]	50.72 51.02 [38.07,65.53]	52.15 51.94 [39.24,65.39]
Modelo 5	192.60 191.66 [134.73,255.27]	161.23 161.06 [124.91,199.75]	220.21 220.31 [152.95,294.78]	224.82 223.02 [156.35,304.73]
Modelo 6	104.94 105.24 [79.50,129.94]	108.56 108.88 [85.64,135.19]	152.44 151.57 [109.15,201.52]	102.50 102.68 [79.09,127.66]

Tabla 5.20: ν -SVM. Inclusión de variables dicotómicas y línea de base en términos de MAE empírico y remuestreo bootstrap para los diferentes métodos y para los productos de la categoría de leche. En negrita se indica para cada producto el método con menor MAE.

A continuación se muestra una comparación usando *bootstrap* que permite tomar decisiones sobre la inclusión de la variable dicotómica y la línea de base. En las Tablas 5.21, 5.22, ?? y 5.24 correspondientes a los métodos k-NN, GRNN, MLP con salidas individuales y ν -SVM, respectivamente. Para cada una de las modelizaciones anteriores, y para cada una de los 6 productos de leche, se ha realizado un test *bootstrap* con tres estadísticos diferentes (Δ MAE, Δ IC, Δ IC_{sup}) para comprobar la existencia de diferencias significativas en unas hipótesis dadas. En concreto, para el resto del experimento se han analizado tres hipótesis diferentes cuya nomenclatura es:

- **Hipótesis 1.** H_0 : la inclusión de la línea de base no es individualmente significativa.
- **Hipótesis 2.** H_0 : la inclusión de la variable dicotómica no es individualmente significativa.
- **Hipótesis 3.** H_0 : la inclusión de la línea de base y la variable dicotómica no son conjuntamente significativas.

A continuación se comprueban estas hipótesis para cada uno de los métodos y para cada uno de los 6 productos de la categoría de leche.

Si se considera como H_0 la Hipótesis 1, en la primera columna de las Tablas 5.21, 5.22, 5.23 y 5.24, correspondientes a k-NN, GRNN, MLP y ν -SVM, respectivamente, se observa que las prestaciones de todos los métodos mejoran cuando se incluye la estacionalidad que presenta el Modelo 6 (véase Figura 4.3(f)), y es preferible la inclusión de la línea de base como variable de entrada. Además, todos los métodos menos GRNN, mejoran si para estimar el número de unidades del Modelo 5 (PulevaClasica) se incluye la línea de base como variable de entrada, como se observa en la Figura 4.3(e). Para los datos del Modelo 4, si se observan los resultados obtenidos del test bootstrap, en ningún método se obtienen diferencias significativas que nos permitan concluir si es preferible la inclusión de la línea de base. El Modelo 3, correspondiente a la marca del distribuidor, y como se observa en la Figura 4.3(c) puede observar una ligera tendencia, la cual es detectada por k-NN. Sin embargo, con GRNN se obtiene que es mejor no incluir la línea de base. Para el análisis del Modelo 2, representado en la Figura 4.3(b), sólo se observan diferencias significativas para GRNN y ν -SVM, siendo mejor no incluir la línea de base. Por último, ν -SVM es el único método que obtiene diferencias significativas para el Modelo 1, y se obtiene que es preferible la inclusión de la línea de base.

Los resultados de considerar la Hipótesis 2 como la hipótesis nula se muestran en la segunda columna de las Tablas 5.21, 5.22, 5.23 y 5.24. Para los métodos k-NN, MLP y ν -SVM, las diferencias significativas resultantes de realizar un test bootstrap pareado demuestran que es preferible incluir la variable dicotómica como entrada del método. En concreto, con k-NN se

obtienen diferencias significativas en el Modelo 3, con MLP en el Modelo 1 y con ν -SVM en los Modelos 1, 3 y 5. Cuando se considera GRNN como método de aprendizaje, el MAE es estadísticamente mayor cuando se incluye la variable dicotómica en los Modelos 2, 4 y 5.

Los resultados de realizar un remuestreo bootstrap pareado para comprobar la inclusión de la línea de base y la variable dicotómica cuando se consideran como métodos de aprendizaje estadístico k-NN, GRNN, MLP y ν -SVM se recogen respectivamente en la tercera columna de las Tablas 5.21, 5.22, 5.23 y 5.24. Para los métodos k-NN, MLP y ν -SVM la inclusión de la variable dicotómica no cambia el análisis, por ello, se puede decir, que para estos métodos el comportamiento de la Hipótesis 1 y 3 son similares. Cuando se considera MLP como método de aprendizaje estadístico, se había obtenido que era preferible incluir la línea de base para estimar el número de unidades vendidas, sin embargo, la inclusión de la línea de base y la variable dicotómica no son conjuntamente significativas, se obtiene menor MAE cuando no se incluye ninguna de estas dos variables.

Como conclusiones de este experimento se puede decir que no existe un comportamiento consistente para todo los métodos. No obstante, se pueden extraer algunos resultados. Por ejemplo, para el Modelo 1 se comprueba que el mejor resultado se obtiene cuando se incluye la variable dicotómica. En el Modelo 2, del análisis de todos los métodos se extrae que es mejor no incluir la línea de base como variable del modelo, pero existen resultados opuestos sobre la inclusión de la variable dicotómica. En el Modelo 3 es complicado extraer conclusiones sobre la inclusión de la línea de base y la variable dicotómica, ya que los resultados obtenidos depende del método usado. Si se observan los resultados obtenidos en el Modelo 4 se puede concluir que es mejor no incluir ni la línea de base ni la variable dicotómica de forma individual, además, cuando se modeliza con GRNN se obtiene que los resultados obtenidos empeoran cuando se incluyen conjuntamente la línea de base y la variable dicotómica como variables de entrada del modelo. Por último, cabe resaltar que el Modelo 6 es el más consistente de todos, las estimaciones obtenidas tiene menor MAE cuando se modelizan los modelos incluyendo como variables de entrada tanto la línea de base como la variable dicotómica.

	SiLBSiVD vs SiVDNoLB	SiVDSiLB vs NoVDSiLB	SiVBSiLB vs NoVDNoLB
Modelo 1	12.58 [-3.57,38.38] -0.17 [-117.46,57.24] 3.67 [-111.23,57.24]	0.26 [-14.77,15.13] 25.91 [-39.46,107.44] 27.22 [-83.61,107.43]	12.49 [-3.99, 38.16] 1.91 [-111.23,57.24] 2.05 [-111.23,57.24]
Modelo 2	12.48 [-3.69,31.84] 8.52 [-15.89,39.00] 8.54 [-11.66,39.00]	5.23 [-8.79,20.85] -13.57 [-42.45,0.00] -14.12 [-38.29,0.00]	10.66 [-8.79,35.43] -3.37 [-27.15,1.73] -3.04 [-27.15,1.73]
Modelo 3	4.57 [-2.71,12.41] -14.78 [-57.50,-0.00] -11.64 [-42.25,-0.00]	1.18 [-8.38,10.63] -20.27 [-73.89,-0.00] -15.40 [-62.35,-0.00]	0.74 [-4.41,6.06] -4.08 [-14.59,-0.00] -1.51 [-6.28,-0.00]
Modelo 4	1.32 [-0.97,3.89] -1.48 [-8.36,0.00] 0.32 [-0.00,0.00]	0.57 [-2.32, 3.53] -1.71 [-7.77,0.00] 0.04 [-0.00,1.33]	1.34 [-0.89, 3.66] -1.63 [-8.36,2.50] 0.31 [-0.00,0.50]
Modelo 5	-59.31 [-115.51,-17.09] -241.95 [-488.21,-54.48] -245.35 [-488.21,-56.55]	1.20 [-11.59,14.48] -13.44 [-73.93,39.63] -18.26 [-73.93,37.68]	-60.66 [-121.03,-14.57] -244.80 [-496.22,-0.01] -255.53 [-509.33,-21.54]
Modelo 6	-47.20 [-87.94,-11.10] -350.01 [-508.58,-45.71] -354.44 [-508.58,-75.33]	8.26 [-6.15,23.69] 9.55 [-8.14,74.77] 8.32 [-6.94,73.95]	-53.44 [-98.44,-17.74] -312.01 [-435.94,-45.71] -317.11 [-437.02,-44.15]

Tabla 5.21: *k*-NN. Leche. Bootstrap pareado para comprobar la inclusión de la línea de base y la variable dicotómica con: ΔMAE , ΔIC y ΔIC_{sup} . Diferencias significativas en negrita.

	SiLBSiVD vs SiVDNoLB	SiVDSiLB vs NoVDSiLB	SiVDSiLB vs NoVDNoLB
Modelo 1	-53.43 [-147.93,37.80] -223.78 [-497.98,141.59] -204.32 [-473.60,166.68]	-37.51 [-86.86,11.14] -161.54 [-293.76,33.44] -179.28 [-284.58,-11.04]	-72.20 [-144.57,5.77] -64.44 [-420.75,295.53] -74.84 [-434.34,276.46]
Modelo 2	-43.89 [-95.38,3.77] -188.09 [-582.37,68.43] -207.32 [-605.42,5.81]	10.68 [-12.16,34.08] -20.69 [-67.19,52.68] -21.16 [-54.46,51.27]	-8.08 [-45.91,25.83] -144.43 [-387.30,25.73] -141.60 [-385.43,7.28]
Modelo 3	4.69 [-13.67,22.08] 27.26 [-41.09,82.78] 17.09 [-45.40,71.84]	0.47 [-4.61,5.59] -10.53 [-43.51,8.75] -7.25 [-40.50,7.13]	-2.54 [-9.61,4.35] -17.95 [-81.90,10.33] -16.25 [-79.13,8.07]
Modelo 4	8.19 [-7.27,23.57] -46.42 [-94.89,33.56] -41.29 [-81.36,39.65]	3.62 [-1.04,8.52] -9.54 [-22.32,0.02] -2.94 [-9.82,3.07]	-3.15 [-12.43,8.76] -2.86 [-20.36,9.11] -0.49 [-17.12,5.46]
Modelo 5	-93.37 [-145.91,-46.58] -233.97 [-631.45,121.13] -243.19 [-643.04,82.00]	10.17 [-0.51,21.31] 54.60 [-63.18,131.18] 59.74 [-55.54,132.96]	4.79 [0.76,11.41] -232.26 [-458.21,99.52] -216.93 [-453.73,104.00]
Modelo 6	-68.49 [-113.36,-24.13] -223.63 [-333.56,-82.72] -226.32 [-323.73,-79.45]	5.32 [-4.10,15.57] 32.14 [-26.08,107.40] 36.98 [-14.16,111.83]	-46.52 [-99.77,-0.24] -259.64 [-391.46,-72.73] -255.81 [-377.29,-67.09]

Tabla 5.23: MLP. Leche. Bootstrap pareado para comprobar la inclusión de la línea de base y la variable dicotómica con: ΔMAE , ΔIC y ΔIC_{sup} . Diferencias significativas en negrita.

	SiLBSiVD vs SiVDNoLB	SiVDSiLB vs NoVDSiLB	SiVBSiLB vs NoVDNoLB
Modelo 1	78.53 [-5.75,215.65] 825.77 [-38.94,2376.33] 736.16 [-38.94,2376.33]	56.12 [-0.82,145.20] 559.54 [-0.01,1562.99] 561.59 [-0.00,1563.00]	79.32 [-8.32,216.43] 847.96 [-38.94,2376.33] 856.12 [-38.94,2376.33]
Modelo 2	73.11 [-1.08,181.56] 760.36 [0.00,1433.00] 799.93 [0.00,1450.00]	75.53 [0.00,193.23] 702.90 [-0.01,1395.00] 726.38 [-0.01,1395.00]	72.07 [-1.51,177.45] 746.05 [0.00,1450.00] 768.29 [0.00,1433.00]
Modelo 3	50.00 [1.00,127.62] 386.85 [0.00,840.00] 407.40 [0.00,840.00]	54.58 [-0.55,132.50] 394.40 [-4.33,821.00] 377.88 [0.00,821.00]	52.28 [1.14,129.71] 404.03 [0.00,840.00] 393.19 [0.00,840.00]
Modelo 4	18.72 [0.21,45.09] 115.91 [0.00,256.00] 113.01 [0.00,256.00]	20.86 [0.48,51.31] 142.55 [-0.00,275.25] 138.85 [0.00,275.00]	18.72 [0.28,44.31] 115.97 [0.00,256.00] 113.47 [0.00,256.00]
Modelo 5	31.10 [-35.42,113.40] 545.49 [-417.00,1383.00] 489.91 [-417.00,1383.00]	100.43 [0.52,254.93] 1138.02 [-0.00,2248.40] 1131.48 [-0.00,2248.40]	31.15 [-41.89,121.03] 530.19 [-417.00,1383.00] 550.69 [-417.00,1383.00]
Modelo 6	-33.61 [-67.43,-7.46] -162.36 [-450.94,28.87] -174.87 [-452.23,14.84]	-3.05 [-23.56,18.32] -5.63 [-67.43,94.41] -5.01 [-66.08,95.77]	-34.55 [-69.53,-6.65] -161.94 [-453.78,28.87] -173.00 [-469.17,14.84]

Tabla 5.22: *GRNN. Leche*. Bootstrap pareado para comprobar la inclusión de la línea de base y la variable dicotómica con: ΔMAE , ΔIC y ΔIC_{sup} . Diferencias significativas en negrita.

	SiLBSiVD vs SiVDNoLB	SiVDSiLB vs NoVDSiLB	SiVBSiLB vs NoVDNoLB
Modelo 1	-44.79 [-90.70,-9.15] -18.49 [-175.90,64.90] -45.98 [-182.30,23.60]	-77.59 [-203.48,-4.23] -637.45 [-2544.40,92.20] -629.26 [-2529.60,73.00]	-27.62 [-57.47,0.37] 6.43 [-72.30,117.30] -6.18 [-56.70,112.90]
Modelo 2	1.35 [-16.33,20.06] 60.09 [1.23,144.06] 71.01 [-0.49,141.23]	3.17 [-3.77,9.81] 6.10 [-22.44,38.04] 7.97 [-19.80,33.49]	0.77 [-16.91,17.39] 51.92 [9.58,90.25] 54.54 [17.59,83.58]
Modelo 3	-5.67 [-18.32,6.30] -18.65 [-54.75,33.66] -16.95 [-50.25,36.91]	-14.35 [-29.58,-1.45] -70.88 [-169.62,8.99] -72.78 [-166.82,8.99]	-0.70 [-16.35,12.77] -18.25 [-53.41,23.82] -17.23 [-51.19,24.78]
Modelo 4	15.77 [-0.07,46.28] 139.70 [-7.79,531.42] 135.78 [-2.31,543.53]	10.99 [-0.86,35.63] 125.89 [-8.69,462.69] 124.35 [-2.26,462.80]	17.52 [0.36,47.84] 124.89 [-8.91,523.58] 135.10 [-7.40,543.52]
Modelo 5	-81.67 [-134.42,-34.95] -294.96 [-429.30,-69.00] -306.17 [-460.60,-63.10]	-14.89 [-29.58,-2.38] -6.93 [-60.88,91.57] -7.12 [-57.91,87.96]	-69.42 [-119.73,-22.66] -298.26 [-444.00,-62.70] -302.06 [-445.20,-63.90]
Modelo 6	-43.85 [-96.08,-2.32] -311.67 [-496.09,-10.70] -320.81 [-508.25,-22.87]	7.48 [-8.18,27.49] 11.34 [-47.18,111.56] 10.12 [-46.15,111.88]	-43.57 [-87.91,-4.99] -306.34 [-518.65,-17.55] -314.97 [-518.97,-26.12]

Tabla 5.24: ν -SVM. *Leche*. Bootstrap pareado para comprobar la inclusión de la línea de base y la variable dicotómica con: ΔMAE , ΔIC y ΔIC_{sup} . Diferencias significativas en negrita.

Conclusiones y líneas futuras

En este capítulo se presentan las conclusiones del proyecto. Se hace una distinción entre el estudio del comportamiento de los diferentes productos a las promociones realizadas y los diferentes métodos de estimación. Por último, se comentan posibles líneas de trabajo futuras, distinguiendo entre los métodos de regresión, el comportamiento de los productos promocionados y la importancia de la marca del distribuidor.

6.1. Conclusiones

El objetivo específico de este trabajo ha sido proponer un procedimiento operativo para el diagnóstico de modelo y la selección de variables para métodos de aprendizaje estadístico en aplicaciones de efectividad promocional.

En cuanto a las bases de datos analizadas, se ha trabajado con datos reales de un año procedentes de un distribuidor minorista, lo cual implica dos consecuencias diferenciadas. Por una parte, durante todo el análisis realizado se debe tener presente que los datos están fuertemente relacionados con el comportamiento del consumidor final, surgiendo así, resultados bastantes complejos para ser explicados a través del uso de métodos empíricos simples. Por otra, el hecho de trabajar con una base de datos real, y no sintética, lleva asociada la posibilidad de encontrar más datos atípicos o dependientes tanto del entorno interno (como promociones realizadas en la misma categoría de productos o en otras categorías de productos, entre otras) como externo (por ejemplo, promociones realizadas por los competidores). Éstas pueden ser algunas de las causas que expliquen los resultados obtenidos en el Capítulo 4, donde se observan las relaciones directas existentes entre los índices de precios y los volúmenes de ventas obtenidos para los diferentes

productos analizados de la categoría de leche y cerveza.

En cuanto a los resultados del modelado promocional incluyendo efectos cruzados, del primer experimento se concluye que es complicado decidir cuál de los 4 métodos de modelado basado en aprendizaje (k-NN, GRNN, MLP y SVM) presenta mejores prestaciones en términos de la observación directa del MAE. Por ello, se utiliza un test estadístico claramente definido y operativo, basado en remuestreo *bootstrap*. De esta aproximación se puede extraer diferentes conclusiones. En primer lugar, para los modelos de la categoría de leche, se obtiene que las prestaciones de la SVM son mejores que las de k-NN y GRNN; si se comparan estos dos métodos, son mejores las prestaciones obtenidas con k-NN. Para los datos de la categoría de cerveza, las prestaciones obtenidas considerando SVM como método para diseñar el modelo de respuesta promocional son mejores que las obtenidas con GRNN. Sin embargo, cuando para esta categoría de producto se contrastan las prestaciones de SVM vs k-NN, para algunos modelos son mejores las prestaciones de la SVM, mientras que para otros, las de k-NN.

Debido a las estimaciones obtenidas con SVM, se ha comprobado si los resultados mejoraban con la inclusión de un núcleo semiparamétrico, de tal forma que las variables dicotómicas se modelen con un núcleo paramétrico. Con los datos disponibles, se puede concluir que no existen diferencias significativas entre ambos análisis.

Tras realizar el diseño del MLP con salidas individuales (Experimento 4), se obtiene que para los modelos de la categoría de leche, MLP con salida conjunta ofrece mejores prestaciones que con salidas individuales; sin embargo, para la categoría de cerveza es preferible estimar el número de unidades vendidas considerando MLP con salidas individuales.

Del análisis conjunto de los experimentos realizados en el proyecto, se puede concluir que el remuestreo *bootstrap* permite contrastar estadísticamente las prestaciones de los diferentes métodos analizados. No obstante, el diagnóstico y selección de variables están condicionados por las características del método promocional analizado, así como por las características propias del producto, por lo que no se puede extraer una conclusión global para ambas categorías de producto. Por tanto, es recomendable no confiar en un único método a la hora de construir estimaciones promocionales a partir de datos reales.

6.2. Líneas futuras

Las promociones de ventas se están convirtiendo hoy en día en una extensa área de investigación, debido a la cantidad de información existente para su análisis, así como al auge de las técnicas de aprendizaje máquina, las cuales ofrecen buenos resultados tanto para estimación

como para predicción. En el presente proyecto, se ha analizado el uso de diferentes métodos de aprendizaje estadístico así como la selección de variables. De este modo, el trabajo aquí presentado supone un punto de partida para futuras investigaciones que puedan orientarse hacia la mejora del mismo o directamente hacia su puesta en funcionamiento. A continuación se destacan algunas posibles líneas de investigación:

- **Análisis en diferentes bases de datos.** En el presente proyecto se han analizado las promociones realizadas en los dos productos de alimentación, en concreto, bebidas. Se propone el análisis en otras categorías de productos, tanto de alimentación como de droguería o perfumería, entre otras.
- **Análisis de promociones en un distribuidor mayorista.** Se propone el análisis del comportamiento de promociones realizadas por un distribuidor mayorista, las cuales se dirigen a un intermediario y no al consumidor final. Además, se propone la comparación con las promociones realizadas por un distribuidor minorista.
- **Efectos cruzados de las promociones sobre las ventas de otras categorías de producto.** Los descuentos aplicados en determinados productos dentro una categoría pueden influir en las ventas obtenidas en otra categoría de producto. Se propone un análisis tanto de la canibalización experimentada, así como de los resultados obtenidos en función del producto o de la marca que se ponga en promoción.
- **Inclusión de variables adicionales que aporten información sobre el entorno en el que ocurre la promoción.** La información tanto del entorno interno como externo puede ser crucial a la hora de comprobar si la promoción realizada ha supuesto un aumento del número de unidades vendidas.

Comparativa de figuras de mérito

En este apartado se analiza la variación existente en la salida obtenida en función de la medida de prestación elegida. Se realizan dos análisis diferentes, por un lado, se muestran los valores predichos en función de la medida de prestación elegida para encontrar los parámetros libres, por otro, se analizan la variación existente entre la salida obtenida y la salida real usando tres medidas de prestación distintas, Willmott's D, RMAE y RMSE. Los resultados obtenidos se recogen en las Tablas [A.1](#), [A.2](#) y [A.3](#). Para cada modelo (Modelo 1, Modelo 2, Modelo 3, Modelo 4, Modelo 5 y Modelo 6) y para los diferentes métodos estudiados (k-NN, GRNN, GRNNind, MLP, MLPind y SVM) se muestra la salida obtenida habiendo elegido los parámetros libres en función de la D de Willmott (columna 1), MAE (columna 2) y MSE (columna 3). En cada tabla aparece la variación existente entre las observaciones realizadas y las obtenidas usando como medida de prestación D de Willmott (Tabla [A.1](#)), RMAE (Tabla [A.2](#)) y RMSE (Tabla [A.3](#)).

Con los resultados obtenidos en las tres tablas anteriores se han realizado dos nuevos análisis. Por un lado, se ha comprobado si existe variación estadística en la salida obtenida en función de la medida de prestación elegida para elegir los parámetros elegidos. Por otro lado, y en función de las conclusiones extraídas del análisis previo, se ha comparado si son significativamente iguales las salidas obtenidas con los diferentes métodos.

En la Tabla [A.4](#) se verifica que la salida obtenida no es significativamente dependiente de la medida de prestación usada para elegir los parámetros libres óptimos. Por ello, y en consonancia con el desarrollo del presente proyecto, se comparan los diferentes métodos usando las salidas obtenidas habiendo elegido los parámetros libres con MAE.

En el análisis de la Tablas [A.5](#) y [A.6](#) cabe destacar en primer lugar la existencia de un nuevo método, GRNN con salidas inviduales, se elige una anchura, σ , diferente para cada modelo, y

DWillmott				
		DWillmott	MAE	MSE
Modelo 1	k-NN	0.5739	0.5723	0.5546
	GRNN	0.4183	0.0744	0.1066
	GRNNInd	0.5004	0.5004	0.5004
	MLP	0.7431	0.7389	0.7125
	MLPind	0.7088	0.7025	0.7031
	SVM	0.5644	0.4567	0.5634
Modelo 2	k-NN	0.6725	0.6534	0.6325
	GRNN	0.6859	0.6946	0.7048
	GRNNInd	0.7169	0.6708	0.7048
	MLP	0.7185	0.7193	0.7005
	MLPind	0.6522	0.6736	0.6555
	SVM	0.7476	0.7170	0.7385
Modelo 3	k-NN	0.1913	0.1452	-0.0452
	GRNN	0.1708	0.3502	0.3554
	GRNNInd	0.5261	0.2934	0.2934
	MLP	0.2414	0.2319	0.3449
	MLPind	0.4102	0.3722	0.3524
	SVM	0.4510	0.3503	0.3496
Modelo 4	k-NN	0.6873	0.6869	0.6757
	GRNN	0.6700	0.2347	0.2511
	GRNNInd	0.2344	0.5467	0.6184
	MLP	0.5572	0.5558	0.5155
	MLPind	0.4684	0.4423	0.4308
	SVM	0.7023	0.6991	0.6991
Modelo 5	k-NN	0.9316	0.9268	0.9213
	GRNN	0.9334	0.9256	0.9266
	GRNNInd	0.9336	0.9933	0.9935
	MLP	0.9238	0.9252	0.9231
	MLPind	0.9171	0.9184	0.9175
	SVM	0.9447	0.8777	0.9092
Modelo 6	k-NN	0.9562	0.9562	0.9542
	GRNN	0.9574	0.9605	0.9608
	GRNNInd	0.9611	0.9868	0.9873
	MLP	0.9435	0.9452	0.9320
	MLPind	0.9282	0.9270	0.9300
	SVM	0.9611	0.9539	0.9607

Tabla A.1: Cálculo de la DWillmott habiendo obtenido los parámetros libres en función de: a) DWillmott; b) MAE c) MSE. En negrita se remarcan los valores con mayor DWillmott.

RMAE				
		DWillmott	MAE	MSE
Modelo 1	k-NN	0.1468	0.1467	0.1479
	GRNN	0.1569	0.1671	0.1675
	GRNNInd	0.1556	0.1556	0.1556
	MLP	0.1332	0.1323	0.1382
	MLPind	0.1398	0.1419	0.1430
	SVM	0.1327	0.1333	0.1327
Modelo 2	k-NN	0.13501	0.1289	0.1299
	GRNN	0.1300	0.1189	0.1146
	GRNNInd	0.1171	0.1133	0.1146
	MLP	0.1307	0.1299	0.1264
	MLPind	0.1241	0.1278	0.1283
	SVM	0.1120	0.1061	0.1133
Modelo 3	k-NN	0.1348	0.1315	0.1636
	GRNN	0.1345	0.1189	0.1196
	GRNNInd	0.1543	0.1155	0.1155
	MLP	0.1252	0.1602	0.1147
	MLPind	0.1155	0.1162	0.1182
	SVM	0.1163	0.1107	0.1108
Modelo 4	k-NN	0.1681	0.1680	0.1725
	GRNN	0.1719	0.1765	0.1757
	GRNNInd	0.5518	0.1622	0.1642
	MLP	0.1939	0.1912	0.1955
	MLPind	0.1792	0.1837	0.1846
	SVM	0.1401	0.1364	0.1364
Modelo 5	k-NN	0.1049	0.0985	0.1021
	GRNN	0.1079	0.1170	0.1163
	GRNNInd	0.1077	0.0384	0.0387
	MLP	0.1207	0.1620	0.1147
	MLPind	0.1294	0.1303	0.1266
	SVM	0.1048	0.1101	0.1232
Modelo 6	k-NN	0.12581	0.1258	0.1290
	GRNN	0.1291	0.1437	0.1405
	GRNNInd	0.1336	0.0768	0.0762
	MLP	0.1699	0.1689	0.1815
	MLPind	0.1900	0.1854	0.1916
	SVM	0.1382	0.1521	0.1367

Tabla A.2: Cálculo RMAE habiendo obtenido los parámetros libre en función de: a) DWillmott; b) MAE; c) MSE. En negrita se remarcan los valores con menor RMAE.

		RMSE		
		DWillmott	MAE	MSE
Modelo 1	k-NN	0.1894	0.1896	0.1894
	GRNN	0.2019	0.2136	0.2122
	GRNNInd	0.2007	0.2007	0.2007
	MLP	0.1654	0.1676	1.1718
	MLPind	0.1813	0.1833	0.1847
	SVM	0.1819	0.1858	0.1819
Modelo 2	k-NN	0.1609	0.1579	0.1580
	GRNN	0.1582	0.1480	0.1466
	GRNNInd	0.1470	0.1530	0.1466
	MLP	0.1605	0.1599	0.1574
	MLPind	0.1584	0.1632	0.1645
	SVM	0.1410	0.1411	0.1412
Modelo 3	k-NN	0.1696	0.1660	0.2046
	GRNN	0.1642	0.1462	1.1464
	GRNNInd	0.2060	0.1452	0.1452
	MLP	0.1597	0.1264	0.1478
	MLPind	0.1500	0.1551	0.1565
	SVM	0.1420	0.1405	0.1406
Modelo 4	k-NN	0.2090	0.2092	0.2092
	GRNN	0.2117	0.2180	0.2171
	GRNNInd	0.7598	0.2001	0.1983
	MLP	0.2246	0.2229	0.2290
	MLPind	0.2243	0.2266	0.229
	SVM	0.1779	0.1748	0.1660
Modelo 5	k-NN	0.1288	0.1281	0.1266
	GRNN	0.1328	0.1410	0.1389
	GRNNInd	0.1332	0.0528	0.0518
	MLP	0.1621	0.1182	0.1509
	MLPind	0.2216	0.2107	0.2089
	SVM	0.1253	0.1480	0.1568
Modelo 6	k-NN	0.1573	0.1573	0.1587
	GRNN	0.1722	0.2049	0.2002
	GRNNInd	0.1908	0.1609	0.1035
	MLP	0.2406	0.2390	0.2605
	MLPind	0.3902	0.3886	0.3918
	SVM	0.2008	0.2602	0.2012

Tabla A.3: Cálculo *RMAE* habiendo obtenido los parámetros libre en función de: a) *DWillmott*; b) *MAE*; c) *MSE*. En negrita se remarcan los valores con menor *RMSE*.

Modelos	Métodos	DWill - MAE	MSE - MAE	MSE - DWill
Modelo 1	k-NN	0.97 [-5.24, 9.38]	1.15 [-4.82, 8.67]	0.00 [0.00,0.00]
	GRNN	-9.90 [-42.10, 21.7]	0.40 [-3.40, 4.40]	0.60 [-18.00, 38.90]
	MLP	0.90 [-4.10, 6.70]	6.10 [-3.10, 14.70]	5.00 [-2.70,0.01]
	SVM	0.90 [-5.00,9.3]	0.10 [-0.04,0.80]	0.00 [0.00,0.00]
Modelo 2	k-NN	6.10 [-3.20, 16.80]	3.00 [-0.3, 1.00]	-0.0055 [-0.0156, 0.0034]
	GRNN	16.40 [-2.50, 36.40]	0.6 [-0.0021, 0.0038]	-0.0153 [-0.0323, 0.0005]
	MLP	0.80 [-0.80, 2.80]	-0.0035 [-0.0107, 0.0039]	-0.0042 [-0.0114,0.0030]
	SVM	6.1 [-30.00,16.00]	0.0003 [-0.0003, 0.0010]	-0.0055 [-0.0156,0.0034]
Modelo 3	k-NN	0.0032 [-0.0022, 0.0097]	0 [0,0]	-0.0031 [-0.0098, 0.0024]
	GRNN	0.0154 [0.0042,0.0260]	0.0006 [-0.0006, 0.0020]	-0.0150 [-0.0254, -0.0041]
	MLP	-0.0012 [-0.0031,0.0003]	-0.0115 [-0.0249, 0.0013]	-0.0100 [-0.0233,0.0029]
	SVM	0.0032 [-0.002,0.009]	0 [0, 0]	-0.0031 [-0.0098, 0.0024]
Modelo 4	k-NN	0.0001 [-0.0014, 0.0011]	0 [0,0]	-0.0001 [-0.0011, 0.0015]
	GRNN	-0.0050 [-0.0449, 0.0337]	-0.0009 [-0.0026, 0.0011]	0.0037 [-0.0365, 0.0433]
	MLP	0.0026 [-0.0015,0.0089]	0.0042 [-0.0030, 0.0115]	0.0015 [-0.0076,0.0109]
	SVM	0.0001 [-0.0014, 0.0011]	0 [0, 0]	-0.0001 [-0.0011, 0.0015]
Modelo 5	k-NN	0.0061 [-0.0017, 0.0160]	0.0024 [-0.0026, 0.0082]	-0.0039 [-0.0083, -0.0001]
	GRNN	-0.0091 [-0.0310, 0.0122]	-0.0007 [-0.0035, 0.0022]	0.0085 [-0.0102, 0.0273]
	MLP	0.0025 [-0.0015, 0.0090]	-0.0037 [-0.0166, 0.0082]	-0.0059 [-0.0187,0.0062]
	SVM	0.0061 [-0.0017,0.0160]	0.0024 [-0.0026, 0.0082]	-0.0039 [-0.0083, -0.0001]
Modelo 6	k-NN	0 [0,0]	0.0008 [-0.0021, 0.0044]	0.0009 [-0.0022, 0.0043]
	GRNN	-0.0149 [-0.0487, 0.0130]	-0.0032 [-0.0069, -0.0000]	0.0119 [-0.0126, 0.0406]
	MLP	0.0011[-0.0018,0.0043]	0.0125 [-0.0120, 0.0408]	0.0115 [-0.0142,0.0391]
	SVM	0 [0,0]	0.0008 [-0.0021, 0.0044]	0.0009 [-0.0022, 0.0043]

Tabla A.4: *Media e intervalo de confianza obtenido de realizar un remuestreo bootstrap para elegir qué estadístico (DWill, MAE y MSE) presenta mejores prestaciones para elegir los parámetros libres en los diferentes modelos y para los diferentes métodos.*

no una común como se había elegido hasta el momento. Además, se realiza la comparación en métodos relativos, D de Willmott y RMAE, y no en términos absolutos como se muestra en el Experimento 2.

MÉTODOS	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
SVM-k-NN	-0.0994 [-0.20,0.06]	0.0696 [-0.02,0.20]	0.2594 [-0.06,0.92]	0.0256 [-0.06,0.18]	-0.0514 [-0.14,0.00]	-0.0012 [-0.02,0.02]
SVM-GRNN	-0.0313 [-0.17,0.17]	0.0475 [-0.00,0.10]	0.0544 [-0.04,0.15]	0.1592 [0.04,0.33]	0.8491 [0.73,0.92]	0.8705 [0.83,0.91]
SVM-GRNNi	-0.0316 [-0.16,0.17]	0.0475 [0.00,0.10]	0.0560 [-0.03,0.16]	0.1613 [0.04,0.34]	0.8506 [0.73,0.92]	0.8704 [0.83,0.91]
SVM-MLP	-0.2829 [-0.53,-0.07]	0.0182 [-0.14,0.25]	0.1424 [-0.11,0.53]	0.1491 [0.04,0.29]	-0.0483 [-0.14,0.02]	0.0097 [-0.01,0.03]
SVM-MLPi	-0.2432 [-0.49,-0.00]	0.0670 [-0.13,0.35]	-0.0149 [-0.30,0.27]	0.2568 [0.10,0.42]	-0.0435 [-0.17,0.05]	0.0290 [-0.01,0.10]
k-NN-GRNN	0.4812 [0.19,0.74]	-0.0479 [-0.18,0.04]	-0.2455 [-0.76,0.03]	0.4377 [0.17,0.64]	0.0009 [-0.02,0.02]	-0.0049 [-0.03,0.01]
k-NN-GRNNi	0.0694 [-0.02,0.17]	-0.0229 [-0.18,0.09]	-0.1951 [-0.77,0.09]	0.1355 [-0.01,0.26]	0.9014 [0.85,0.93]	0.8715 [0.83,0.90]
k-NN-MLP	-0.1784 [-0.50,0.04]	-0.0499 [-0.23,0.18]	-0.1089 [-0.67,0.28]	0.1260 [-0.03,0.25]	0.0114 [-0.03,0.03]	0.0012 [-0.01,0.03]
k-NN-MLPi	-0.1435 [-0.51,0.11]	-0.0014 [-0.20,0.27]	-0.2712 [-0.94,0.15]	0.2381 [-0.01,0.43]	0.0101 [-0.05,0.08]	0.0301 [-0.02,0.10]
GRNN-GRNNi	-0.4082 [-0.68,-0.12]	0.0225 [-0.01,0.06]	0.0472 [-0.06,0.10]	-0.3037 [-0.45,-0.12]	0.9013 [0.85,0.93]	0.8764 [0.83,0.91]
GRNN-MLP	-0.6593 [-1.10,-0.33]	-0.0075 [-0.18,0.21]	0.1311 [-0.12,0.49]	-0.3121 [-0.53,-0.05]	-0.0002 [-0.03,0.03]	0.0160 [0.00,0.03]
GRNN-MLPi	-0.6174 [-1.09,-0.25]	0.0441 [-0.15,0.35]	-0.0341 [-0.36,0.27]	-0.2059 [-0.48,0.04]	0.0076 [-0.05,0.08]	0.0362 [-0.00,0.11]
GRNNi-MLP	-0.2498 [-0.57,-0.02]	0.0261 [-0.22,0.23]	0.0850 [-0.16,0.49]	-0.0105 [-0.13,0.09]	-0.9006 [-0.94,-0.86]	-0.8611 [-0.90,-0.82]
GRNNi-MLPi	-0.2150 [-0.57,0.04]	0.0229 [-0.19,0.32]	-0.0800 [-0.35,0.20]	0.0960 [-0.09,0.26]	-0.8937 [-0.94,-0.82]	-0.8414 [-0.90,-0.76]
MLP-MLPi	0.0373 [-0.01,0.14]	0.0486 [-0.08,0.21]	-0.1586 [-0.55,0.09]	0.1097 [-0.05,0.27]	0.0086 [-0.04,0.08]	0.0189 [-0.02,0.09]

Tabla A.5: Comparativa bootstrap pareada de los métodos analizados en este proyecto para cada modelo usando como estadístico *DWillmott*. En negrita se marcan las diferencias estadísticamente significativas.

MÉTODOS	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
SVM-k-NN	-0.0133 [-0.04,0.01]	-0.0228 [-0.04,-0.00]	-0.0208 [-0.04,-0.00]	-0.0310 [-0.06,-0.00]	0.0117 [-0.01,0.04]	0.0251 [-0.03,0.09]
SVM-GRNN	-0.0334 [-0.06,-0.01]	-0.0077 [-0.02,-0.00]	-0.0081 [-0.02,-0.00]	-0.0402 [-0.07,-0.01]	-0.0070 [-0.03,0.02]	0.0086 [-0.02,0.05]
SVM-GRNNi	-0.0222 [-0.05,0.00]	-0.0071 [-0.01,0.00]	-0.0048 [-0.01,-0.00]	-0.0261 [-0.05,-0.01]	-0.6595 [-0.70,-0.62]	-0.4000 [-0.51,-0.27]
SVM-MLP	0.0011 [-0.04,0.04]	-0.0238 [-0.05,0.00]	-0.0155 [-0.03,0.00]	-0.0544 [-0.08,-0.03]	-0.0080 [-0.05,0.03]	-0.0162 [-0.06,0.02]
SVM-MLPi	-0.0082 [-0.04,0.03]	-0.0224 [-0.05,0.00]	-0.0055 [-0.02,0.01]	-0.0471 [-0.08,-0.02]	-0.0211 [-0.07,0.02]	-0.0335 [-0.12,0.03]
k-NN-GRNN	-0.0197 [-0.05,0.01]	0.0154 [-0.00,0.03]	0.0126 [-0.00,0.03]	-0.0093 [-0.05,0.03]	-0.0189 [-0.05,0.01]	-0.0176 [-0.06,0.02]
k-NN-GRNNi	-0.0091 [-0.03,0.01]	0.0159 [-0.01,0.04]	0.0157 [-0.00,0.03]	0.0059 [-0.02,0.04]	-0.6715 [-0.70,-0.64]	-0.4242 [-0.50,-0.35]
k-NN-MLP	0.0139 [-0.02,0.05]	-0.00073 [-0.03,0.03]	0.0050 [-0.02,0.03]	-0.0223 [-0.06,0.01]	-0.0198 [-0.06,0.01]	-0.0430 [-0.10,0.01]
k-NN-MLPi	0.0046 [-0.03,0.04]	0.00087 [-0.03,0.03]	0.0149 [-0.01,0.04]	-0.0164 [-0.06,0.03]	-0.0320 [-0.09,0.01]	-0.0590 [-0.18,0.02]
GRNN-GRNNi	0.0112 [-0.02,0.05]	0.0006 [-0.00,0.01]	0.0033 [-0.00,0.01]	0.0145 [-0.01,0.04]	-0.6533 [-0.69,-0.62]	-0.4068 [-0.49,-0.31]
GRNN-MLP	0.0350 [-0.00,0.07]	-0.0158 [-0.04,0.01]	-0.0074 [-0.02,0.01]	-0.0152 [-0.05,0.02]	-0.0009 [-0.03,0.03]	-0.0252 [-0.06,0.01]
GRNN-MLPi	0.0250 [-0.01,0.06]	-0.0142 [-0.04,0.01]	0.0030 [-0.02,0.02]	-0.0078 [-0.04,0.02]	-0.0134 [-0.07,0.03]	-0.0418 [-0.16,0.03]
GRNNi-MLP	0.0228 [-0.01,0.06]	-0.0164 [-0.05,0.01]	-0.0108 [-0.03,0.01]	-0.0293 [-0.05,-0.01]	0.6523 [0.61,0.69]	0.3836 [0.28,0.48]
GRNNi-MLPi	0.0136 [-0.02,0.05]	-0.0147 [-0.04,0.01]	-0.0009 [-0.02,0.02]	-0.0222 [-0.05,0.01]	0.6395 [0.57,0.69]	0.3661 [0.20,0.49]
MLP-MLPi	-0.0094 [-0.03,0.00]	0.0023 [-0.02,0.03]	0.0103 [-0.00,0.02]	0.0072 [-0.02,0.04]	-0.0120 [-0.07,0.03]	-0.0188 [-0.11,0.04]

Tabla A.6: Comparativa bootstrap pareada de los métodos analizados en este proyecto para cada modelo usando como estadístico RMAE. En negrita se marcan las diferencias estadísticamente significativas.

APÉNDICE **B**

Presupuesto del proyecto

El objetivo de este Ápendice es calcular el coste asociado a la realización del presente proyecto. Para ello, se realiza una estimación tanto del coste de los medios materiales como de los recursos humanos empleados.

B.1. Coste de los medios materiales

En la Tabla **B.1** se muestra el coste asociado al material utilizado para realizar el proyecto. Al coste total de los medios materiales hay que añadirle el coste del emplazamiento del trabajo donde se ha realizado el proyecto, así como los costes asociados al alumbrado, calefacción, aire acondicionado, servicios de limpieza, muebles de oficina, tomas de corriente y conexiones de red de banda ancha. Si el coste aproximado de estos servicios asciende a 600 €/mes, y se suma los costes aproximados de desplazamiento y dietas, 150 €/mes, el coste total obtenido es de 750 €/mes. Teniendo en cuenta que se ha estimado la duración de trabajo del proyecto en 12 meses, el coste total asciende a 9000 €.

EL coste total de los medios materiales asciende a 11060, resultado de sumar el coste total del material más los costes de emplazamiento.

B.2. Coste del personal

Para estimar el coste del personal se tiene en cuenta que el Colegio Oficial de Ingenieros de Telecomunicación estima el sueldo medio de un ingeniero de telecomunicaciones en 30000 € brutos

Concepto	Coste €
Ordenador personal	800
Licencia Windows 7 Professional Full	310
Licencia Microsoft Office 2010 Home & Student PC	150
Licencia MATLAB	500
Impresora + tóner + papel	150
Gasto de oficina	50
Material bibliográfico	100
Coste total	2060

Tabla B.1: Coste del material.

anuales. Sin embargo, el coste de un ingeniero recién titulado es inferior, ronda los 15000 € brutos anuales. Por tanto, el coste del personal asciende a 15000 €.

B.3. Coste de dirección

Cuando se realiza un proyecto de ingeniería, de forma general, se estima el salario del director del proyecto como el 7% de la suma del coste material más el coste de personal. Por tanto, el coste de dirección del proyecto asciende a 1824.2 €.

B.4. Coste total

En la Tabla B.4 se desglosa los diferentes costes que suman el coste total del presente trabajo. Para evitar imprevistos o cambios en la planificación, se añade un coste adicional que varía entre el 1% y el 5% del coste total del proyecto.

Concepto	Coste €
Coste de los materiales y suministros	11060
Coste de personal	15000
Coste de dirección	1824.2
Coste añadido (reserva)	2606
Coste total del proyecto	30490.2

Tabla B.2: Coste total del proyecto.

Bibliografía

- [1] J. Quelch. How to market in a recession. *Marketing KnowHow*, September 2008.
- [2] M.P. Martínez and A.I. Jiménez. Evolución y tendencias de la marca de distribuidor en los mercados de gran consumo: el caso de España. *Cuadernos de gestión*, 9(2), 2009.
- [3] O. González. ¿Qué busca el comprador de marcas de distribuidor? Caracterización del comprador de marca de distribuidor por beneficios buscados. *Universia Business Review*, 24:76–94, 2009.
- [4] M.P. Martínez. *La influencia de las características del descuento de precio promocional y de la marca en el incremento de las ventas: aplicación de la metodología SVM con datos de escáner minorista*. PhD thesis, Departamento de Economía y Empresa Área de Comercialización e Investigación de Mercados. Universidad de Castilla La Mancha, 2004.
- [5] F.J. Villalba and P.C. Iñaki. La promoción de ventas en los mercados de consumo: Propuesta de un nuevo enfoque de gestión promocional. *Cuadernos de Gestión*, 2(2), 2002.
- [6] T.M. Mitchell. *Machine learning*. McGraw-Hill Higher Education, 1997.
- [7] H.J. Van Heerde, P.S.H. Leeflang, and D.R. Wittink. Building models for marketing decision: past, present and future. *SOM theme F: Interactions between consumers and firms*, 2000.
- [8] D.R. Bell, J. Chiang, and V. Padamanabhan. The decomposition of promotional response: an empirical generalization. *Marketing Science*, 18(4):504–546, 1999.
- [9] P.S.H. Leeflang and D.R. Wittink. Building models for marketing decisions: past, present and future. *International Journal of Research in Marketing*, 17:178–185, 2000.

- [10] R.G. Blattberg, R. Briescha, and E.J. Fox. How promotion work. *Marketing Science*, 14:122–132, 1995.
- [11] A.M. Alonso. Un ejemplo de bootstrap suavizado. *Lecturas Matemáticas*, 23:11–24, 2002.
- [12] F.J.De La Ballina. *La Investigación de promoción de ventas en España*. Netbiblo. Serie Finanzas y Empresa, 2007.
- [13] E. Bigné. *Promoción comercial*. ESIC Editorial, 2003.
- [14] L. Fischer and J. Espejo. *Mercadotecnia*. McGraw-Hill, 1993.
- [15] P. Kotler and G. Armstrong. *Fundamentos del marketing*. Prentice Hall, 1994.
- [16] J. Rivera and M.D. Juan. *La promoción de ventas variable clave del marketing*. ESIC Editorial, 2000.
- [17] W.J. Stanton, M.J. Etzel, and B.J. Walker. *Fundamentos de marketing*. McGraw-Hill, 1992.
- [18] B. Rabassa. *Promoción de ventas. Cómo se prepara una campaña*. Pirámide, 1987.
- [19] M.P. Martínez, A. Mollá, and J.L. Rojo. Using daily store-level data to understand price promotion effects in a semiparametric regression model. *Retailing and Consumer Services*, 3(13):193–204, 2006.
- [20] M.P. Martínez, A. Mollá, and J.L. Rojo. Assessing the impact of temporary retail price discounts intervals using SVM semiparametric regression. *The International Review of Retail, Distribution and Consumer Research*, 16(2):181–197, 2006b.
- [21] Y.Yustas. *Análisis del Efecto de las Promociones sobre las Ventas en Establecimientos Detallistas de Alimentación*. PhD thesis, Universidad Pontificia Comillas de Madrid, 2000.
- [22] H.J. Van Heerde, P.S.H. Leeflang, and D.R. Wittink. Semiparametric analysis to estimate the deal effect curve. *Journal of Marketing Research*, 38(2):197–215, May 2001.
- [23] R.G. Blattberg and S. Neslin. *Sales Promotions Models*. Number 553-609. Research and Management Science, north-holland edition, 1993. Handbook in Operations.
- [24] V.R Nijs, M.G. Dekimpe, J.B. Steenkamps, and D.M. Hanssesns. The category-demand effects of price promotions. *Marketing Science*, 1(1-22), 20.
- [25] M.P. Martínez, A. Mollá, and J.L. Rojo. La influencia de las características del descuento de precio promocional minorista sobre las ventas: Aplicación de un modelo semiparamétrico. *Revista Española de Investigación de Marketing. ESIC.*, 9(2):159–181, 2005.

- [26] R.G. Walters. Retail promotions and retail store performance. a test of some key hypotheses. *Journal of Retailing*, 64(2):153–180, 1988.
- [27] K. Jedidi, C.F. Mela, and S. Gupta. Managing advertising and promotion for long-run profitability. *Marketing Science*, 18(1):1–22, 1999.
- [28] K.L. Ailawadi and B. Harlam. An empirical analysis of the determinants of retail margins: The role of store-brand share. *Journal of Marketing*, 68(1):147–165, 2004.
- [29] S. Srinivasan, K.Pauwels, D.M. Hanssens, and M.G. Dekimpe. Do promotions benefit manufacturers, retailers, or both? *Management Science*, 50(5):617–629, 2004.
- [30] K. Ailawadi, S. Neslin, and K. Gedenk. Pursuing the value-conscious consumer: store brands versus national brand promotions. *Journal of Marketing*, 65(January,):71–89, 2001.
- [31] S.F. Chen, K.B. Monroe, and Y.C. Lou. The effects of framing price promotion messages on consumers’ perceptions and purchase intentions. *Journal of Retailing*, 74(3):353–372, 1998.
- [32] E. Martínez and T. Montaner. The effect of consumer’s psychographic variables upon deal-proneness. *Journal of Retailing and Consumer Services*, 13(3):157–168, 2006.
- [33] S. Burton, D. Lichtenstein, and R. Netemeyer. Exposure to sales flyers and increased purchases in retail supermarkets. *Journal of Advertising Research*, 39(5):7–14, 1999.
- [34] M. Miranda and L. Kónya. Directing store flyers to the appropriate audience. *Journal of Retailing and Consumer Service*, 14(3):175–181, 2007.
- [35] M. Schmidt and M. Bjerre. Can recipients of sales flyers be segmented? *International Journal of Advertising*, 22(375-391.), 2003.
- [36] J.P. Molina. Estimación de la curva de respuesta promocional mediante máquinas de vectores soporte. Master’s thesis, Universidad Carlos III de Madrid. Escuela Politécnica Superior, 2007.
- [37] E.A. Nadaraya. *Nonparametric estimation of probability densities and regression curves*, volume 20. Mathematics and its Applications (Soviet Series), Publishers Group, 1989.
- [38] C. Bishop. *Neural networks for pattern recognition*. Oxford University Press, 1995.
- [39] L. Devroye, L.Gyorfi, and G.Lugosi. *Probabilistic theory of pattern recognition. Stochastic modelling and applied probability*. Springer, 1996.

-
- [40] F. Moreno. *Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano*. PhD thesis, Universidad de Alicante. Departamento de Lenguajes y Sistemas Informáticos, 2004.
- [41] B.M. del Brío and A.S.Molina. *Redes neuronales y sistemas borrosos*. Ra-Ma, 1997.
- [42] P.D. Wasserman. *Advanced Methods in Neural Computing*. Van Nostrand Reinhold, 1993.
- [43] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [44] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 1999.
- [45] K. Vapnik. *The nature of statistical learning theory*. Springer-Verlang, 1995.
- [46] B. Schölkopf and A. J. Smola. Learning with kernels. *MIT Press*, 2002.
- [47] S. Haykin. *Neural networks. A comprehensive foundation*. Prentice-Hall, 1999.
- [48] C.J. Willmott. On the validation of models. *Phys Geogr*, 2:184–194, 1981.
- [49] S. Geisser and W.F. Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74, 1979.
- [50] J.L. Rojo. Clasificación de arritmias cardiacas mediante vectores soporte. Master's thesis, Dpto. de señales, sistema y radiocomunicaciones. E.T.S.I. de Telecomunicaciones. Univ. Politécnica de Madrid., 2000.
- [51] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap*. Chapman & Hall, 1997.