Introducción a la Estadística

Victoria Ruiz, Iván Ramírez, Emanuele Schiavi

FUNDAMENTOS MATEMÁTICOS

CURSO 2023-2024

UNIVERSIDAD REY JUAN CARLOS

MÁSTER EN VISIÓN ARTIFICIAL

ÍNDICE

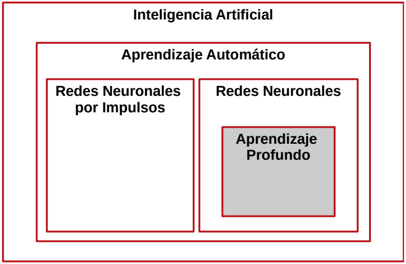
Section 1

Introducción

Teoría de la Probabilidad vs Teoría de la Información vs Aprendizaje Profundo

- La **teoría de la probabilidad** nos proporciona un medio de cuantificar la incertidumbre.
  - Conozco el sistema, pero no tengo recursos.
  - El sistema está construido.
- La **teoría de la información** se puede cuantificar la cantidad de incertidumbre que hay en una distribución de probabilidad.
- El **aprendizaje automático** usa las herramientas anteriores para construir algoritmos que realicen una tarea concreta a partir de los datos de nuestro problema.

Teoría de la Probabilidad vs Teoría de la Información vs Aprendizaje Profundo



Section 2

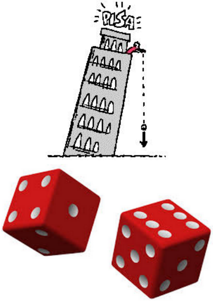
Fundamentos de Estadística Bayesiana.

Probabilidad frecuentista vs Probabilidad Bayesiana

- **Probabilidad Frecuentista**: La probabilidad de un determinado suceso será la frecuencia relativa de este después de que el experimento se haya repetido un número grande de veces.
  - No siempre puedo realizar un experimento varias veces, como los ensayos clínicos.
- **Probabilidad Bayesiana**: Utiliza información *a priori*. Es decir, se utiliza el conocimiento previo que se tiene antes de realizar el experimento.
  - Esta información puede cambiar en función de quién o cómo se realiza el experimento. Si no se tiene información se supondrá que todos los eventos tienen la misma probabilidad de ocurrir

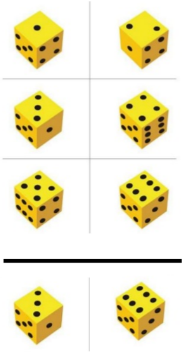
Experimento, espacio muestral, muestra y suceso

- Para obtener la respuesta a nuestra pregunta, tendré que realizar un **experimento**. Un experimento puede ser:
    - **Experimento determinista**.
    - **Experimento aleatorio**:
        - Se conocen todos sus posibles resultados.
        - No puede predecirse resultado.
        - Puede haber distintos resultados.



Experimento, espacio muestral, muestra y suceso

- El conjunto de resultados posibles de dicho experimento es lo que se conoce como **espacio muestral**, S.
- Los resultados concretos de realizar el experimento se denominan **muestras**.
- Se denomina **suceso** asociado al experimento E a cualquier subconjunto, $A \subseteq S$, del espacio muestral.



Ejercicios: ¿Cómo se representaría el suceso obtener un número par? ¿y obtener un número primo?

Probabilidad Condicionada

- En muchas situaciones, aun sin conocer el resultado del experimento, sí que disponemos de información adicional, y dicha información debe incorporarse al cálculo de la probabilidad de los sucesos.
- Es decir, nos interesa la **probabilidad condicional**. Denotamos la probabilidad condicional de $Y = y$ dado $X = x$ como $P(Y = y|X = x)$. Esto se puede calcular como:

$$P(Y = y|X = x) = \frac{P(Y = y \cap X = x)}{P(X = x)} \tag{1}$$

- La probabilidad condicional sólo se define cuando $P(X = x) > 0$, ya que en caso contrario no es posible calcularla.

Probabilidad Condicionada

- Cualquier distribución de probabilidad conjunta sobra varias variables puede descomponerse en la distribución condicional sobre una de las variables. Es decir, se cumple la **regla de la cadena de la probabilidad**. Por ejemplo,

$$\begin{aligned} & P(X_1 = x_1, \cdots, X_n = x_n) = \\ = \quad & P(X_1 = x_1) \cdot \prod_{i=2}^{n} P(X_i = x_i | X_1 = x_1, \cdots X_{i-1} = x_{i-1}) \end{aligned}$$

Independencia

- Se dice que un conjunto de variables aleatorias es **independiente** si y sólo si la función de distribución conjunta es producto de las funciones de distribución marginales. Es decir,

$$\forall x_1 \in X_1, \cdots, x_n \in X_n,$$
$$P(X_1 = x_1, \cdots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$$

- Un conjunto de variables $X_1, \cdots, X_n$ son condicionalmente independientes de otro conjunto $Y_1, \cdots, Y_m$, si la distribución de probabilidad condicionada sobre $X_1, \cdots, X_n$ factoriza de la siguiente manera con los valores de $Y_1, \cdots, Y_m$:

$$\forall x_1 \in X_1, \cdots, x_n \in X_n, y_1 \in Y_1, \cdots y_n \in Y_m,$$
$$P(X_1 = x_1, \cdots, X_n = x_n | Y_1, \cdots, Y_m) =$$
$$P(X_1 = x_1 | Y_1, \cdots, Y_m) \cdot \ldots \cdot P(X_n = x_n | Y_1, \cdots, Y_m)$$

Independencia y Probabilidad Condicionada

- Dos sucesos A y B, asociados a un experimento aleatorio, son independientes si el hecho de que ocurra uno de ellos no influye en la probabilidad de que ocurra el otro; es decir, si

$$P(A|B) = P(A) \quad \text{o lo que es igual} \quad P(B|A) = P(B)$$

- En otro caso, se dice que los sucesos son dependientes. La independencia de dos sucesos se puede caracterizar por la siguiente propiedad:
  Los sucesos A y B son independientes si y sólo si

$$P(A \cap B) = P(A) \cdot P(B)$$

Regla de Bayes

- Si conocemos $P(Y|X)$ y $P(X)$, pero necesitamos saber $P(X|Y)$ se puede calcular gracias a la **Regla de Bayes**:

$$P(X|Y) = \frac{P(X)P(Y|X)}{P(Y)}$$

donde P(Y) se puede calcular como

$$P(Y) = \sum_x P(Y|x)P(x)$$

- Si tenemos los sucesos $A_1, A_2, \cdots A_k$, de manera que todos tienen probabilidad positiva $(P(A_i) > 0 \quad \forall i \in \{1, 2, \cdots, k\})$. Sea B cualquier suceso con $P(B) > 0$. Entonces, se tiene el **Teorema de Bayes**:

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(A_j) \cdot P(B|A_j)}{\sum_{i=1}^{k} (P(A_i) \cdot P(B|A_i))}$$

Variables Aleatorias

- Una **variable aleatoria (v.a)** X es una función que asigna un valor $X(s)$ a cada resultado del experimento $s \in S$.

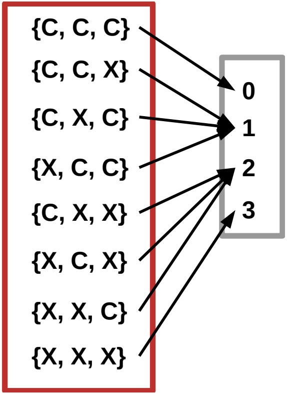
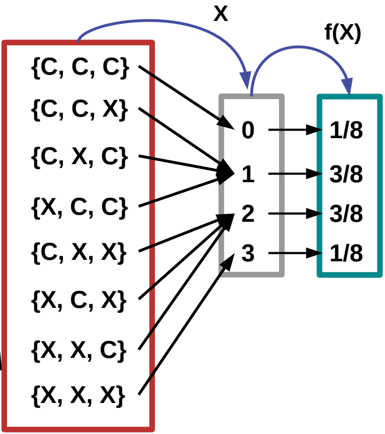


Figura: V.a: Nº de cruces que se han obtenido

Funciones de probabilidad

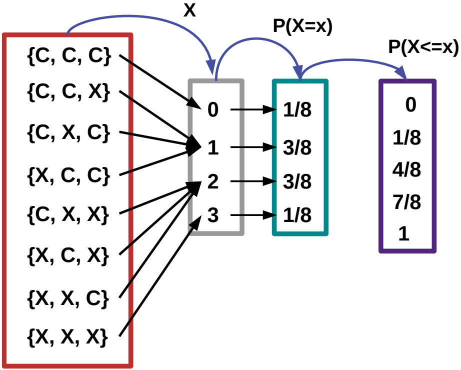
- Para asignar a cada valor una probabilidad de que ocurra, se utilizarán las **funciones de probabilidad**.
- Para determinar la *distribución de probabilidad* de los posibles valores que toma cualquier variable aleatoria X definida sobre S:

$$P(X \in A) = P\{s \in S / X(s) \in A$$

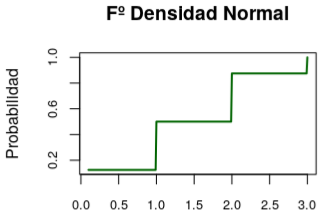


Función de Distribución

- Para asignar a cada suceso la probabilidad de que la variable aleatoria, X, tome un valor igual o inferior a uno dado, x, se utilizará una **función de distribución** $F(x) = P((-\infty, x]) = P(X \leq x)$



Representación gráfica función de masa y de distribución



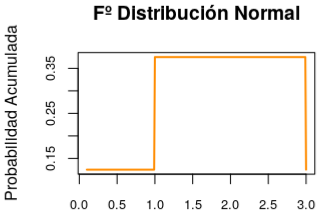


Figura: Función de masa (derecha) y función de distribución (izquierda)

Variables Discretas

- La probabilidad de cualquier suceso $A \subset S$ se calculará como:

$$P(A) = \sum_{s_i \in A} p_i$$

- La función de distribución $F(x)$ de este tipo de variable se calcula como la suma de la función de probabilidad para valores iguales o inferiores al valor x, y resulta por lo tanto una función escalonada:

$$F(x) = \sum_{x_i \leq x} P(x = x_i); \qquad -\infty < x < +\infty$$

Variables Continuas

- Las **variables continuas** son aquellas que toman cualquier valor de los números reales.
- Para asignar a cada posible valor su probabilidad de que ocurra se utiliza la **función de densidad de probabilidad**.

$$F(x) \quad = \quad P(x \in [-, \infty, x])) = \int_{-\infty}^{x} f(y)dy \quad ; -\infty \leq x \leq +\infty$$

- Esta función cumple las propiedades:
  - El dominio I de p debe ser el conjunto de todos los posibles valores de X.
  - $\forall x \in X, f(x) \geq 0$.
  - $\int f(x)dx = 1$. Es decir, la integral de la función de densidad en todo el espacio ha de ser la unidad.

Esperanza de un v.a

- La **esperanza** o valor esperado o media de una variable aleatoria X, denotada por μ o $E[X]$, es una medida central del comportamiento de la variable que representa su centro de masas.
- Se puede interpretar como el *valor medio que tomaría la variable en una larga serie de observaciones*:

$$E[X] = \mu = \begin{cases} \sum_{x_i} x_i p(x_i) & \text{siendo } X_i \text{ variable discreta} \\ \int_{\mathbb{R}} x f(x) dx & \text{siendo } X_i \text{ variable continua} \end{cases}$$

- La esperanza por su definición es un operador lineal, es decir

$$E\left[\sum_{i=1}^{n} c_i X_i\right] = \sum_{i=1}^{n} c_i E[X_i]$$

Varianza de una v.a

- La **varianza** de una variable aleatoria es una medida para determinar el nivel de dispersión que tienen los valores de la variable respecto de su media.
- Se denota por var(X) o σ^2, y se define como la media de las desviaciones a la esperanza al cuadrado.

$$\sigma^2 = E[(X_i - \mu_i)^2] = E[X_i^2] - \mu_i^2$$

- $$V[X] = \sigma^2 = \begin{cases} \sum_{x_i} (x_i - \mu)^2 p(x_i) & \text{siendo } X_i \text{ variable discreta} \\ \int_{\mathbb{R}} (x - \mu)^2 f(x) dx & \text{siendo } X_i \text{ variable continua} \end{cases}$$
- Por definición, la varianza es siempre no negativa y es un operador cuadrático, de modo que

$$var(cX) = c^2 var(X)$$

Desviación Típica y Pseudovarianza

- Obséserve que la varianza va en unidades al cuadrado; para dar una medida relativa a las unidades que se están manejando se utiliza la **desviación típica o estándar**, que se define como la raíz cuadrada de la varianza de la variable, y se denota por σ.
- Durante el curso manejaremos a menudo la **pseudovarianza** que se define como:

$$S_x^2 = \frac{1}{n-1}\sum_{i=1}(x_i - \bar{x})^2$$

Nótese que la varianza y la pseudovarianza se relacionan de la siguiente forma:

$$\sigma_x^2 = \frac{n-1}{n} S_x^2$$

Ejemplo con Matlab

```
LLAMADA
Genero valores de f(x)
Genero 100000 valores de la variable aleatoria:
x=genero_discreta2(100000);
media=mean(x)
pseudovarianza=var(x)
varianza=pseudovarianza*99999/100000
RESULTADO
media = 1.4975
pseudovarianza = 0.7529
varianza = 0.7529
```

Probabilidad Conjunta

- La **función de distribución conjunta** es aquella que asigna a cada conjunto la probabilidad de que las variables aleatorias, $X_1, X_2, \cdots, X_n$, tomen un valor igual o inferior a unos valores dados, $x_1, x_2, \cdots, x_n$. Es decir

$$F_{(X_1,\cdots,X_n)} = P(X_1 \leq x_1, \cdots, X_n \leq x_n)$$

- Análogamente, la **función de probabilidad conjunta** asigna a cada conjunto de sucesos su probabilidad. De este modo se definiá la **función de masa conjunta** como

$$p_{(X_1,\cdots,X_n)}(x_1, \cdots, x_n) = P(X_1 = x_1, \cdots, X_n = x_n)$$

- Y la **función de densidad conjunta** como

$$f_{(X_1,\cdots,X_n)}(x_1, \cdots, x_n)$$

Probabilidad Marginal

- Alguna veces, conocemos la función de distribución sobre el conjunto de variables, pero queremos saber la distribución sobre un subconjunto de ellas, es decir la **distribución marginal de probabilidad**.
- Por ejemplo, supongamos que tenemos las variables discretas X e Y, y conocemos $P(X, Y)$. Se puede conocer P(X) con la "regla de la suma":

$$\forall x \in X, P(X = x) = \sum_y P(X = x, Y = y)$$

- Análogamente, para variables continuas se calcularía como:

$$f(x) = P(X = x) = \int f(x, y) dy$$

Ejm Representación Gráfica v.a Multidimensionales

Con el siguiente código de Matlab se representar gráficamente

```
[x,y] = meshgrid(0:.1:10);
z=exp(-y-x);
f4=figure;
figure(f4);
ax2 = subplot(1,2,1);
mesh(x,y,z)
ax2 = subplot(1,2,2);
z2=(1-exp(-x)).*(1-exp(-y));
mesh(x,y,z2)
```

Ejm Representación Gráfica Función de masa y de Distribución

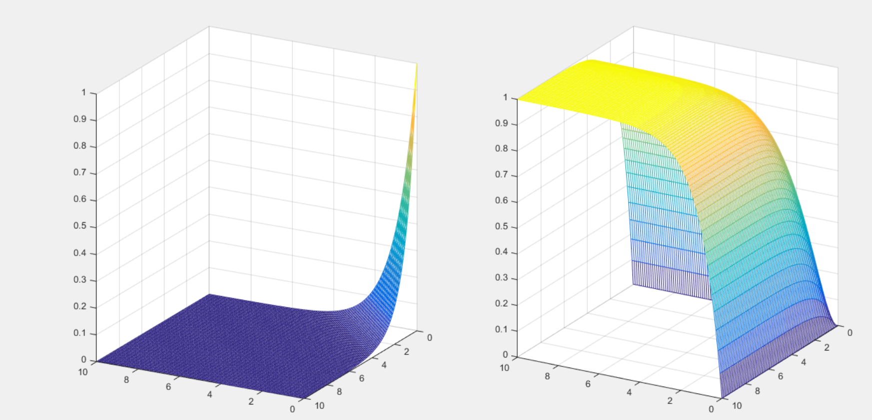


Figura: Función de masa del ejercicio 1.17 (izquierda) y función de distribución

Covarianzas

- Una medida de la dependencia de lineal entre dos variable X_i y X_j viene dada por la **covarianza** (C_{ij} o $\text{cov}(X_i, X_j)$):

$$cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j =)] = E[X_i X_j] - \mu_i \mu_j$$

- Las covarianzas son simétricas, es decir, $C_{ij} = C_{ji}$.
- Si $i = j$, entonces la covarianza es igual a la varianza de la variable.
- Si el valor de la covarianza es positivo, si el valor que se tiene de X_i es mayor que su media μ_i, entonces el valor que se espera para X_j será mayor que su media μ_j.
- Si el valor de la covarianza es negativo, si el valor que se tiene de X_i es mayor que su media μ_i, entonces el valor que se espera para X_j será menor que su media μ_j.

Section 3

Distribuciones

Bernouilli(p)

- Distribución sobre una única variable binaria.
- Se caracteriza por un único parámetro p que representa la probabilidad de que la variable aleatoria tome el valor 1.
- Se utiliza para representar el resultado de experimentos con 2 posibles estados.
- Rango:$\{0, 1\}$
- Media: p.
- Varianza: $p(1-p)$
- Parámetros: $p \equiv$ probabilidad del evento 1.
- Función de masa: $p(x) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \end{cases}$

Binomial(n,p)

- Se define como la suma de n variables de Bernouilli.
- Representa que el experimento estudiado se ha repetido n veces de manera independiente, y se quiere conocer el número de "éxitos".
- Se caracteriza por los parámetro p, que representa la probabilidad de que la variable aleatoria tome el valor 1 y por n, que representa el número de veces que se ha repetido el experimento.
- Rango: $\{0, 1, \cdots, n\}$
- Media: np.
- Varianza: $np(1-p)$
- Parámetros: $p \equiv$ probabilidad del evento ; n, número de veces que se repite el experimento.
- Función de masa: $p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x \in 0, 1, \cdots n \\ 0 & enotrocaso \end{cases}$

Matlab

```
x=0:0.1:10;
Binonimal
parametros
n=10;
p=0.2;
F. de masa Binomial
y_masa_bin=binopdf(x,n,p);
```

```
F. de distribución Binomial
y_dist_bin=binocdf(x,n,p);
estadísticos
[m,v]=binostat(n,p);
generar valores
nrow=1;  numero filas
ncol=10; numero de columnas
x_bino= binornd(n,p,nrow,ncol);
```

Gráficas

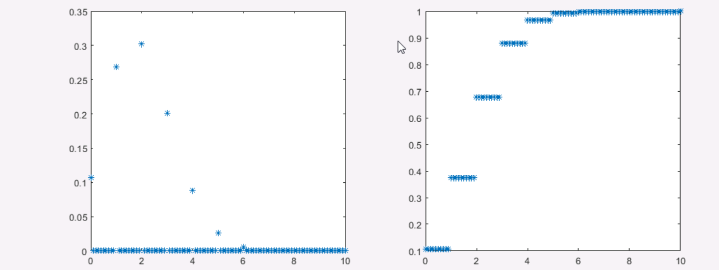


Figura: Función de masa (izquierda) y función de distribución (derecha) de una distribución Binomial(10,0.2)

Geométrica(p)

- Se basa en la distribución de Bernouilli.
- Representa el número de repeticiones necesarias hasta obtener el primer "éxito".
- Esta distribución tiene como característica la falta de memoria.
- Rango: $\{1, 2, \cdots\}$
- Media: $\frac{1}{p}$.
- Varianza: $\frac{1-p}{p^2}$
- Parámetros: $p \equiv$ probabilidad del evento 1.
- Función de masa: $p(x) = \begin{cases} p(1-p)^{x-1} & x \in 1, 2 \cdots \\ 0 & en otro caso \end{cases}$

Matlab

```
x=0:0.1:10;
Geométrica
parámetros
p=0.2;
F. de masa Geométrica
y_masa_geo=geopdf(x,p);
F. de distribución Geométrica
y_dist_geo=geocdf(x,p);
estadísticos
[m,v]=geostat(p);
generar valores
nrow=1; numero filas
ncol=10; numero de columnas
x_geo= geornd(p,nrow,ncol);
```

Gráficas

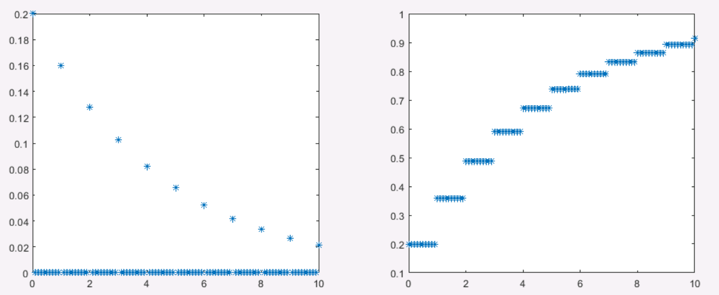


Figura: Función de masa (izquierda) y función de distribución (derecha) de una distribución Geométrica(0.2)

Poisson(L)

- Representa el número de eventos que ocurren en un intervalo de tiempo cuando ocurren con tasa constante
- Rango: $\{0, 1, 2, \cdots\}$
- Media: L.
- Varianza: L
- Parámetros: $L \equiv$ tasa de ocurrencia.
- Función de masa: $p(x) = \begin{cases} \frac{e^{-L}L^x}{x!} & x \in 0, 1, 2 \cdots \\ 0 & enotrocaso \end{cases}$

Matlab

```
x=0:0.1:10;
Poisson:

parámetros
L=5;
F. de masa Poisson')
y_masa_pois=poisspdf(x,L);
F. de distribución Poisson')
y_dist_pois=poisscdf(x,L);
estadísticos
[m,v]=poisstat(L);
generar valores
nrow=1;  numero filas
ncol=10;numero de columnas
x_pois= poissrnd(L,nrow,ncol);
```

Gráficas

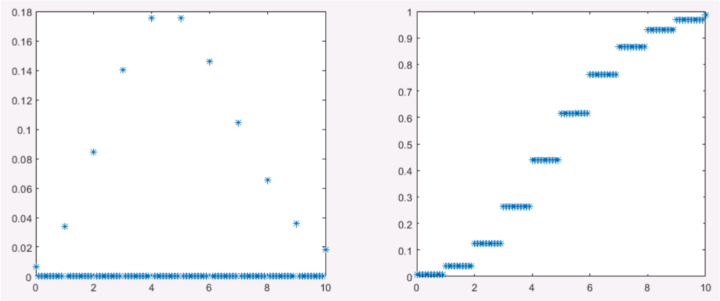


Figura: Función de masa (izquierda) y función de distribución (derecha) de una distribución Poisson(5)

Uniforme(a,b)

- Todos los eventos tienen la misma probablidad.
- Rango: (a, b)
- Media: $\frac{a+b}{2}$.
- Varianza: $\frac{(b-a)^2}{12}$
- Parámetros: a primero valor, y b último valor del intervalo
- Función de densidad: $f(x) = \frac{1}{b-a}$

Matlab

```
x=-3:0.01:3;
Uniforme

parametros
a=-1;
b=1;
F. de densidad Exponencial
y_dens_unif=unifpdf(x,a,b);
F.de distribución Exponencial
y_dist_unif=unifcdf(x,a,b);
estadisticos
[m,v]=unifstat(a,b);
generar valores
nrow=1;  numero filas
ncol=10; numero de columnas
x_unif= unifrnd(a,b,nrow,ncol);
```

Gráficas

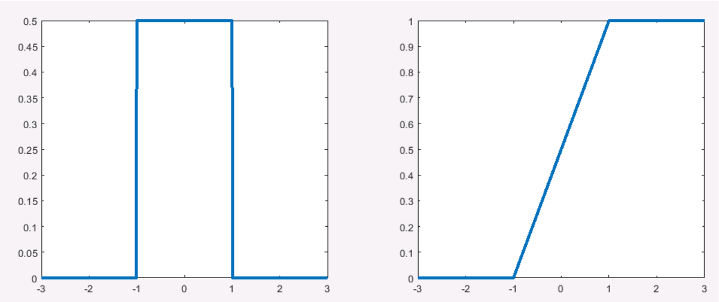


Figura: Función de densidad (izquierda) y función de distribución (derecha) de una distribución Uniforme(-1,1)

Exponencial(λ)

- Se utiliza para obtener tiempo entre sucesos independientes.
- Se caracteriza porque no tiene memoria.
- Rango: $[0, \infty)$
- Media: $\frac{1}{\lambda}$.
- Varianza: $\frac{1}{\lambda^2}$
- Parámetros: $\lambda \equiv$ tasa de ocurrencia.
- Función de densidad: $f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$

Matlab

```
x=-3:0.01:3;
Exponencial

parámetros
lambda=0.5;
F. de densidad Exponencial')

y_dens_exp=exppdf(x,lambda);

F. de dist. Exponencial')
```

```
y_dist_exp=expcdf(x,lambda);


estadísticos
[m,v]=expstat(lambda);
generar valores
nrow=1; numero filas
ncol=10; numero de columnas
x_exp= exprnd(lambda,nrow,ncol);
```

Gráficas

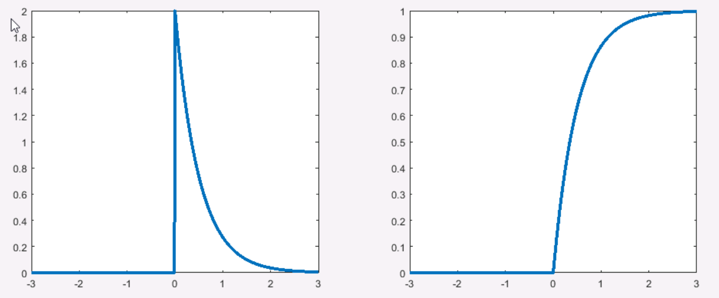


Figura: Función de densidad (izquierda) y función de distribución (derecha) de una distribución Exponencial(0.5)

Normal(μ, σ)

- Es una de la distribuciones más utilizadas y con mayores propiedades.
- Si dos normales son incorreladas, entonces son independientes.
- La combinación lineal de normales es normal.
- Los valores se acumulan en torno a un valor medio (μ), y varían respecto a una desviación típica (σ).
- Rango: $(-\infty, +\infty)$
- Media: μ.
- Varianza: σ^2
- Parámetros: μ, σ.
- Función de densidad: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Matlab

```
x=-3:0.01:3;
Normal
parámetros mu=0; sigma=1;
F. de dens Normal')
y_dens_norm=normpdf(x,mu,sigma);

F. de distribución Normal
y_dist_norm=normcdf(x,mu,sigma);
```

```
estadísticos
[m,v]=normstat(mu,sigma);
generar valores
nrow=1;  numero filas
ncol=10; numero de columnas
x_norm= normrnd(mu,sigma,nrow,ncol
```

Gráficas

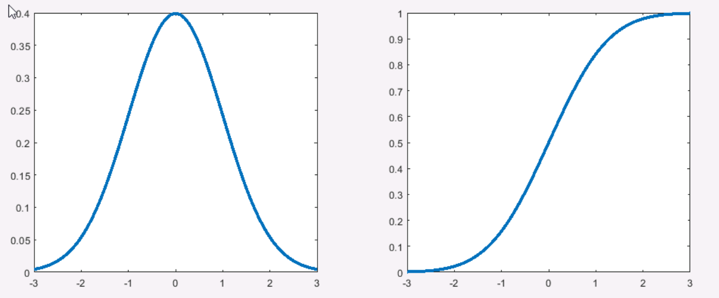


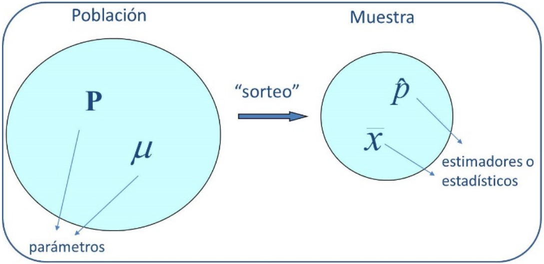
Figura: Función de densidad (izquierda) y función de distribución (derecha) de una distribución Normal(0,1)

Section 4

Estadísticos y Variabilidad

Motivación

- Se quiere conocer una característica de la población (**parámetro poblacional**).
- No se puede o no interesa analizar el conjunto de la **población** $\Rightarrow$ Se trabajará con una **muestra** de la población de estudio, entendiendo por tal una parte representativa de la misma.
- A partir de la muestra se hará una estimación del parámetro poblacional(**estadístico muestral**)que es conocida.



Variabilidad de los estadísticos

Ejemplo:

- Se quiere estimar la distancia media que tienen que viajar los estudiantes de este campus desde que salen de sus casas hasta que llegan al aula.
- D="la distancia que han de viajar los estudiantes de este campus desde su casa hasta el aula"; $D \rightsquigarrow N(\mu, \sigma)$ donde μ y σ^2 son desconocidos.
- Para estimar el parámetro μ podemos escoger una muestra de 10 estudiantes y medir la distancia que viaja cada uno de ellos: 2.43, 3.12, 7.34, 6.33, 2.65, 10.32, 0.55, 6.02, 1.12, 4.33
- Pero también podríamos escoger otra muestra, 4.32, 2.12, 8.65, 4.32, 12.55, 3.02, 1.18, 6.25, 4.44, 8.31
- Las medias serán $\bar{x_1} = 4{,}421$ y $\bar{x_2} = 5{,}516$ respectivamente

Estimadores

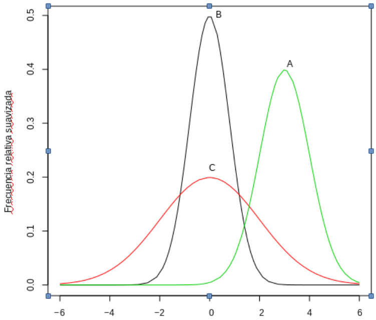
- No todos los estadísticos sirven para estimar los mismos parárametros. Sólo servirán los estadísticos que son **estimadores** de ese parámetro.
- Formalmente, sea X una v.a. con función de densidad $f_\theta(x)$ siendo θ, el parámetro a estimar, llamaremos **estimador** T de θ a cualquier función de la muestra que contiene el parámetro desconocido.

$$\theta = T(x_1, \cdots, x_n)$$

- Como veremos, los estimadores para estimar la media y la varianza más habituales son:
  - **Media Muestral**: $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$
  - **Varianza Muestral**:$S_x^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$
  - **Cuasivarianza Muestral**:$s_x^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2$

Ejemplo

Para estimar la demanda de un cierto producto una empresa dispone de tres especialistas A, B y C. En el siguiente gráfico aparecen histogramas aproximados por densidades, para las diferencias entre las demandas que se han producido en la realidad a lo largo de los últimos 36 meses, y la demanda prevista por cada especialista.



¿Qué especialista podemos decir que se acerca más a la demanda real, y por tanto acierta más en la estimación?

Propiedades de un estimador

Para que un estimador sea bueno es deseable las siguientes **propiedades**:

- Insesgadez: Un estimador es insesgado si la distribución en el muestreo del estimador de θ tiene como esperanza un valor igual a θ, esto quiere decir que la distribución estará centrada alrededor de θ.
- Eficiencia: La varianza del estimador debe ser la menor posible. Entre dos estimadores T1 y T2, si ambos son insesgados, es mejor aquel con menor varianza.
- Consistencia: Al aumentar el tamaño muestral (más datos) mayor debe ser la concentración de la distribución del estimador entorno a θ.

Estimación Puntual

- La **estimación puntual** proporciona la mejor predicción de una cantidad de interés. Esta cantidad suele ser el conjunto de parámetros del modelo, como los pesos en la regresión lineal.
- Denotando por θ el conjunto de parámetros reales y $\hat{\theta}$ el conjunto de parámetros estimados, y sea $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, , \cdots, \mathbf{x}^{(m)}\}$ el conjunto de m datos independientes e idénticamente distribuidos (i.i.d), *estimador puntual* o *estadístico* de cualquier función de datos será:

$$\hat{\theta}_m = g(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, , \cdots, \mathbf{x}^{(m)})$$

Estimador de Máxima Verosimilitud

- Como la función real no se suele conocer, en lugar de adivinar de qué función provienen y calcular después su sesgo y su varianza, nos gustaría tener algún tipo de regla que nos diga qué funciones son buenos estimadores para los diferentes modelos. Esta "regla" se conoce como el *principio de máxima verosimilutd*.
- Intuitivamente, la verosimilitud nos dice *cómo de probable es que los parámentros que he estimado sean correctos a partir de mis datos de entrenamiento*.

Estimador de Máxima Verosimilitud

- Formalmente, consideramos un conjunto de m ejemplos $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(m)}\}$ generados de manera independiente, y de los que desconocemos su distribución f.
- Sea $p_{modelo}(\mathbf{x};\theta)$ una familia de funciones de probilidad sobre el espacio indexado por θ.
- Entonces, el estimador de máxima verosimilitud de θ se define como

$$\theta_{ML} = argmax_\theta\{p_{model}(\mathbf{X};\theta)\} = argmax_\theta \left\{ \prod_{i=1}^{m} p_{model}(\mathbf{x}^{(i)};\theta) \right\}$$

Estimador de Máxima Verosimilud

- Tomando logaritmo, el resultado no cambia y nos transforma el producto en una suma. Por tanto, tenemos que

$$\theta_{ML} = argmax_{\theta} \left\{ \sum_{i=1}^{m} log(p_{model}(\mathbf{x}^{(i)}; \theta)) \right\}$$

- Este estimador se puede generalizar para estimar una probabilidad condicionada $P(\mathbf{Y}|\mathbf{X}, \theta)$:

$$\theta_{ML} = argmax_{\theta} \sum_{i=1}^{m} logP(\mathbf{y}^{(i)}|\mathbf{x}^{(i)}, \theta)$$

Regresión Lineal desde el punto de vista de la Verosimilitud

- Desde el punto de vista de la verosimilitud, en lugar de producir una única predicción $\hat{y}$, pensamos que el modelo lo que realmente está produciendo es la probabilidad condicionada $p(y|\mathbf{x})$.
- Por tanto, el objetivo del algoritmo ahora es ajustar la distribución $p(y|\mathbf{x})$ a todos los posibles valores de y compatibles con $\mathbf{x}$.
- Si suponemos que $p(y|\mathbf{x})$ sigue una distribución normal donde la media viene dada por la función $\hat{y}(\mathbf{x};\omega)$ y de varianza σ^2, como la función de densidad de esa función es:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Regresión Lineal desde el punto de vista de la Verosimilitud

Se tiene que el estimador de máxima verosimilitud viene dado por:

$$\begin{aligned}
\theta_{ML} &= argmax_\theta \left\{ bfx^{(i)}; \theta) \right\} = \\
&= argmax_\theta \left\{ \sum_{i=1}^{m} log \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{||\hat{y}^{(i)} - y^{(i)}||^2)}{2\sigma^2}} \right) = \right\} \\
&= argmax_\theta \left\{ -mlog\sigma - \frac{m}{2} log(2\pi) - \sum_{i=1}^{m} \frac{||\hat{y}^{(i)} - y^{(i)}||^2)}{2\sigma^2} \right\}
\end{aligned}$$

donde $\hat{y}^{(i)}$ es la salida de la regresión del i-ésima entrada $\mathbf{x}^{(i)}$ y m es el número de datos de entrada.

Estadísticos Bayesianos

- Si el problema se enfoca mediante la aproximación Bayesiana, se considera todos los posibles valores de θ para hacer la predicción. Es decir, se tiene un vector de parámetros θ fijo pero desconocido, y se estima $\hat{\theta}$ como una variable aleatoria a partir de los datos.
- Si se tiene alguna información previa de θ, se utilizará para construir la distribución de probabilidad de θ.
- La idea que subyace similar a la del estimador de máxima verosimilitud. Sin embargo, el estimador de máxima verosimilutd hace predicciones usando una estimación de θ, mientras que la aproximación Bayesiana hace predicciones usando una distribución sobre θ.

Estadísticos Bayesianos

Para construir un estadístico Bayesiano:

1. Se escoge distribución a priori. Si no se conoce, se escoge distribución con mucha incertidumbre como la uniforme, donde todos los sucesos tienen la misma proabilidad.
2. Se considera que tenemos un conjunto de datos de muestras $\{x^{(1)}, x^{(2)}, \cdots, x^{(m)}\}$, y se calcula la probabilidad de obtener dicha muestra a partir de θ, $p(x^{(1)}, x^{(2)}, \cdots, x^{(m)}|\theta)$.
3. Lo que realmente nos interesa es la probabilidad de que esos parámetros sean ciertos dados esos datos. De este modo, a partir de la regla de Bayes se puede calcular la probabilidad condicionada o probabilidad a posteriori, $p(\theta|x^{(1)}, x^{(2)}, \cdots, x^{(m)})$:

$$p(\theta|x^{(1)}, x^{(2)}, \cdots, x^{(m)}) = \frac{p(x^{(1)}, x^{(2)}, \cdots, x^{(m)}|\theta)p(\theta)}{p(x^{(1)}, x^{(2)}, \cdots, x^{(m)})}$$

Regresión Lineal desde punto de visa Bayesiano

- Los parámetros del modelo corresponden con el vector ω donde $\hat{y} = \omega^T \mathbf{x}$
- Tomando la distribución normal o gaussiana como la distribucíón de las predicciones, se tiene que:

$$p(\mathbf{y}|\mathbf{X}) \approx exp\left(-\frac{1}{2}\left(\mathbf{y} - \mathbf{X}\omega\right)^T \cdot \left(\mathbf{y} - \mathbf{X}\omega\right)\right)$$

- Suponiendo que la varianza es 1, se obtiene la ecuación del MSE.

Regresión Lineal desde el punto de vista Bayesiano

- Para determinar la probabilidad a posteriori, especificamos la distribución a priori. Tomando de nuevo la distribución normal, tenemos que:

$$p(\omega) \approx exp\left(-\frac{1}{2}\left(\omega - \mu_{\mathbf{0}}\right)^{T} \mathbf{\Lambda_0^{-1}} \left(\omega - \mu_{\mathbf{0}}\right)\right)$$

donde $\mu_{\mathbf{0}}$ y $\mathbf{\Lambda_0}$ son el vector de medias y la matriz de covarianzas. Por tanto, la probabilidad a posteriori será

$$\begin{aligned} p \quad & (\omega|\mathbf{X}, \mathbf{y}) \propto p(\mathbf{y}|\mathbf{X}, \omega) p(\omega) \\ \propto \quad & exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{X}\omega)^{T}(\mathbf{y} - \mathbf{X}\omega)\right) exp\left(-\frac{1}{2}(\omega - \mu_{\mathbf{0}})\Lambda_0^{-1}(\omega - \mu_{\mathbf{0}})\right) \\ \propto \quad & exp\left(-\frac{1}{2}(-2\mathbf{y}^{\mathbf{T}}\mathbf{X}\omega + \omega^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\omega + \omega^{\mathbf{T}}\mathbf{\Lambda_0^{-1}}\omega - \mathbf{2}\mu_{\mathbf{0}}\mathbf{\Lambda_0^{-1}}\omega)\right) \end{aligned}$$

Regresión Lineal desde el punto de vista Bayesiano

- Si definimos ahora $\mathbf{\Lambda_m} = (\mathbf{X^T X} + \mathbf{\Lambda_0^{-1}})^{-1}$ y $\mu_\mathbf{m} = \mathbf{\Lambda_m}(\mathbf{X^T y} + \mathbf{\Lambda_0^{-1}}\mu_\mathbf{0})$, lo anterior se puede reescribir como:

$$
\begin{aligned}
p(\omega|\mathbf{X},\mathbf{y}) \quad &\propto \quad exp\left(-\frac{1}{2}(\omega-\mu_\mathbf{m})\Lambda_m^{-1}(\omega-\mu_\mathbf{m}) + \frac{1}{2}\mu_m^T\Lambda_m^{-1}\mu_m\right) \\
&\propto \quad exp\left(-\frac{1}{2}(\omega-\mu_\mathbf{m})\Lambda_m^{-1}(\omega-\mu_\mathbf{m})\right)
\end{aligned}
$$

- Todos los términos que no contengan los parámetros ω se pueden omitir ya que se consideran constantes y no hay que minimizarlas.

Section 5

Aprendizaje Automático.

Capacidad, Sobre Entrenamiento y Bajo Entrenamiento

- Si el modelo no es capaz de obtener un error suficentemente bajo, se dice que el modelo está *bajoentrenado*.
- Si el error es bueno pero la diferencia entre el error de entrenamiento y el error de validación es muy grande, se dice que el modelo está *sobreentrenado*
- Se puede controlar cuando un modelo es más propenso a estar sobre o bajo entrenado modificando su *capacidad*. Es decir, modificando su habilidad para ajustar funciones. La habilidad de las funciones suele venir por el número de parámetros que contiene.

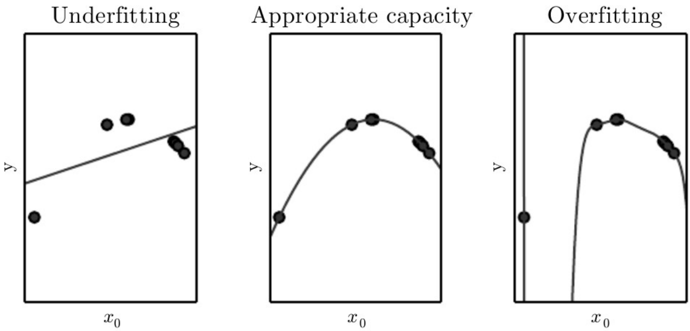


Figura: Ejemplo de capacidad, sobre entrenamiento y bajo entrenamiento

Regresión Lineal desde el punto de vista del Aprendizaje Automático

- Modelo a construir:

$$\hat{y} = \mathbf{w}^T \mathbf{x}$$

- La bondad del modelo se mide por

$$MSE = \frac{1}{m} \sum_i (\hat{y}_i - y_i)^2 = \frac{1}{m} ||\hat{y}_i - y_i||_2^2 g$$

- Para que sea un algoritmo de aprendizaje automático, hay que diseñar un algoritmo que mejore la estimación de los pesos $\mathbf{w}$ de tal forma que se reduzca el MSE $\rightarrow$ minimizar el MSE del conjunto de entrenamiento $\rightarrow$ igualamos a 0 su gradiente.

Regresión Lineal desde el punto de vista del Aprendizaje Automático

Estas ecuaciones se conocen como ecuaciones normales. En la figura 11 se muestra cómo se ajustan los datos y cómo varía el error al entrenar:

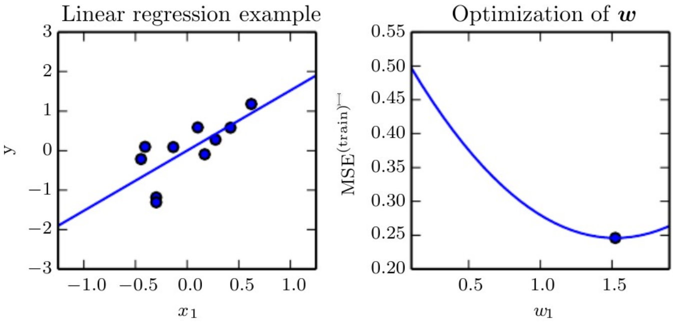


Figura: Ejemplo de regresión lineal y de optimización de ω

Sesgo

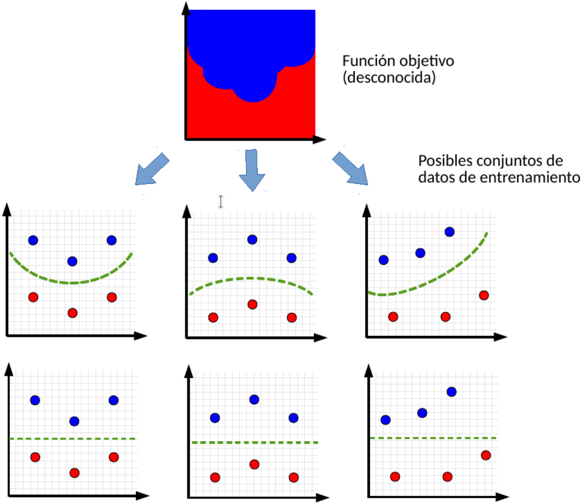
- Formalmente, el **sesgo** (*bias*) de un estimador se define como:

$$bias(\hat{\theta}_m) = E\left[\hat{\theta}_m\right] - \theta_m$$

  Es decir, es la diferencia entre los valores medios que se esperan de los datos estimados y los datos reales.

- La *varianza* ($V(\theta)$) hace referencia a cuánto esperamos que varien los datos como función de los datos de entrada, es decir, cuánto varían los datos en $\hat{f}$. La raíz cuadrada de la varianza se conoce como *error estándar* ($SE(\hat{\theta})$).
- Hay que llegar a un compromiso de sesgo-varianza. Es decir, elegir un modelo que, aun pudiendo mejorar el sesgo y la varianza por separado, ambas medidas sean lo más pequeñas posible.

Ejemplo Sesgo-Varianza



PARTE II - BLOQUE I: Álgebra Lineal y Numérica

Emanuele Schiavi, Iván Ramírez, Victoria Ruiz

FUNDAMENTOS MATEMÁTICOS

CURSO 2023-2024

UNIVERSIDAD REY JUAN CARLOS

MÁSTER EN VISIÓN ARTIFICIAL

Contenido

1. Introducción
   - El álgebra lineal y el procesamiento digital de imágenes.
   - Aplicaciones

Contenido

Contenido

Contenido

Contenido

Contenido

Contenido

Introducción

- El contenido de este primer capítulo es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.

Introducción

- El contenido de este primer capítulo es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.
- El **álgebra lineal** es una rama de las matemáticas que permite codificar, almacenar y operar con las *imágenes digitales*.

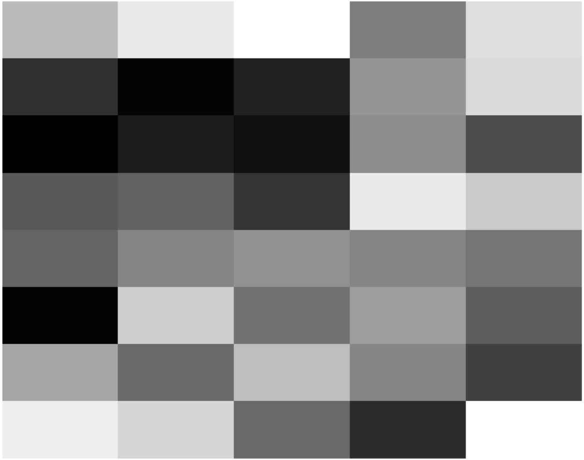
Introducción

- El contenido de este primer capítulo es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.
- El **álgebra lineal** es una rama de las matemáticas que permite codificar, almacenar y operar con las *imágenes digitales*.
- Representación matricial (escalas de grises) y tensorial (escalas de colores) de las imágenes.

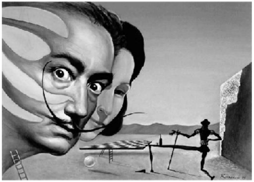
Introducción

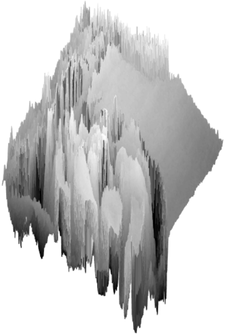
- El contenido de este primer capítulo es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.
- El **álgebra lineal** es una rama de las matemáticas que permite codificar, almacenar y operar con las *imágenes digitales*.
- Representación matricial (escalas de grises) y tensorial (escalas de colores) de las imágenes.
- Cualquier imagen digital, adquirida mediante cámara, se codifica en una estructura algebraica de tipo matriz o tensor.

Introducción



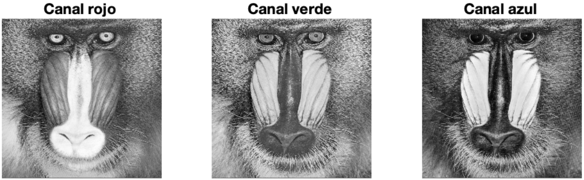
Introducción



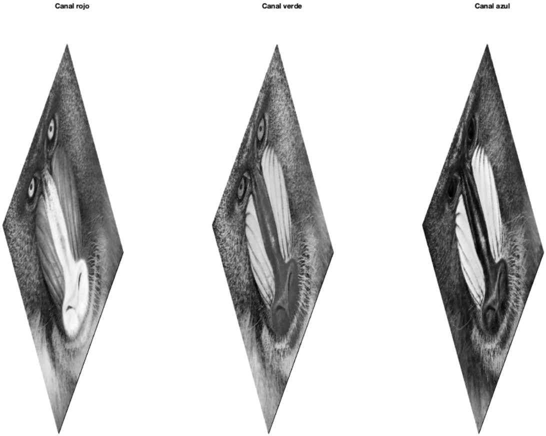


Introducción

Introducción



Introducción



El álgebra lineal y el procesamiento digital de imágenes.

- Eliminación de ruido (Denoising), pérdida de información (Inpainting), Deconvolución y eliminación de emborronamiento (Deblurring), suavizado, super-resolución, segmentación, registro rígido y deformable, etc

El álgebra lineal y el procesamiento digital de imágenes.

- Eliminación de ruido (Denoising), pérdida de información (Inpainting), Deconvolución y eliminación de emborronamiento (Deblurring), suavizado, super-resolución, segmentación, registro rígido y deformable, etc
- El procesamiento digital de imágenes no se limita solamente a retocar o cambiar el tamaño de las imágenes capturadas con la cámara; su uso se extiende a muchos campos de la ciencia y la tecnología.

Aplicaciones

- Medicina, detección remota, transmisión y codificación de datos, robótica, visión artificial, reconocimiento de patrones, industria cinematográfica, procesamiento de imágenes obtenidas de microscopios y restauración y enfocado de imágenes. Superresolución.
- Convoluciones discretas y filtros. La transformada integral de Fourier.
- Transformaciones lineales: traslación, rotación, escalado, reflexión y ajustes de canal, brillo y contraste.

Aplicaciones

- Medicina, detección remota, transmisión y codificación de datos, robótica, visión artificial, reconocimiento de patrones, industria cinematográfica, procesamiento de imágenes obtenidas de microscopios y restauración y enfocado de imágenes. Superresolución.
- Convoluciones discretas y filtros. La transformada integral de Fourier.
- Transformaciones lineales: traslación, rotación, escalado, reflexión y ajustes de canal, brillo y contraste.

Contenido

Dependencia e Independencia lineal

- El concepto de **Dependencia lineal** es fundamental para entender como es posible generar infinitos elementos del espacio vectorial a partir de un número finito de vectores.

Dependencia e Independencia lineal

- El concepto de **Dependencia lineal** es fundamental para entender como es posible generar infinitos elementos del espacio vectorial a partir de un número finito de vectores.
- La generación de vectores se realiza mediante combinaciones lineales de vectores

Combinaciones lineales

Combinaciones lineales

Generación

Formalización

Sea dado un conjunto de m *vectores* $\{\mathbf{v_1}, ..., \mathbf{v_m}\}$ *de* $\mathbb{R}^n$*. Entonces cualquier vector* $\mathbf{w}$ *definido mediante combinación lineal*

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{v}_i = \alpha_1 \mathbf{v}_1 + ... + \alpha_m \mathbf{v}_m, \qquad \alpha_i \in \mathbb{R}$$

pertenece al espacio generado por el conjunto de vectores.

Dependencia e Independencia lineal

Formalización

Diremos que un conjunto de m *vectores* $\{\mathbf{v_1}, ..., \mathbf{v_m}\}$ *de* $\mathbb{R}^n$ *es Linealmente Independiente si el sistema de* n *ecuaciones*

$$\sum_{i=1}^{m} \alpha_i \mathbf{v}_i = \alpha_1 \mathbf{v}_1 + ... + \alpha_m \mathbf{v}_m = \mathbf{0}$$

tiene sólo la solución trivial $\mathbf{0}$ *es decir, todos los coeficientes* α_i *son nulos:*

$$\alpha_1 = ... = \alpha_m = 0$$

- Determinar si un conjunto de vectores es linealmente independiente se reconduce al estudio de un sistema de ecuaciones lineal y homogéneo: $A\mathbf{x} = \mathbf{0}$. El algoritmo de reducción a forma escalonada por filas.

Sistemas de Generadores y Bases. Bases ortogonales

- Los sistemas de generadores de un espacio (o sub-espacio) vectorial permiten generar cualquier elemento del espacio vectorial a partir de un conjunto finito de vectores mediante la operación de combinación lineal.

Sistemas de Generadores y Bases. Bases ortogonales

- Los sistemas de generadores de un espacio (o sub-espacio) vectorial permiten generar cualquier elemento del espacio vectorial a partir de un conjunto finito de vectores mediante la operación de combinación lineal.

Formalización

Sea G un conjunto de m vectores $G = \{\mathbf{v_1}, ..., \mathbf{v_m}\}$ de $\mathbb{R}^n$ siendo $m \leq n$. El sub-espacio $V \subset \mathbb{R}^n$ generado se denota por

$$V = L[\mathbf{v_1}, ..., \mathbf{v_m}] \subset \mathbb{R}^n$$

y los infinitos vectores de V se generan mediante la operación de combinación lineal:

$$V = L[\mathbf{v_1}, ..., \mathbf{v_m}] = \{\mathbf{w} \in \mathbb{R}^n \,/\, \mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{v_i},\ \alpha_i \in \mathbb{R}\} \subset \mathbb{R}^n$$

Sistemas de Generadores y Bases

- Implicitamente un sistema de generadores tiene *redundancia* en la información. La forma de generar un elemento del espacio no es única.

Sistemas de Generadores y Bases

- Implicitamente un sistema de generadores tiene *redundancia* en la información. La forma de generar un elemento del espacio no es única.
- El cálculo del rango de una matriz es una medida de la información presente en la matriz. Es el número máximo de vectores (columnas) linealmente independientes. El concepto de dimensión de un espacio vectorial como número mínimo de generadores.

Sistemas de Generadores y Bases

- Implicitamente un sistema de generadores tiene *redundancia* en la información. La forma de generar un elemento del espacio no es única.
- El cálculo del rango de una matriz es una medida de la información presente en la matriz. Es el número máximo de vectores (columnas) linealmente independientes. El concepto de dimensión de un espacio vectorial como número mínimo de generadores.
- Necesitamos independencia lineal para que un sistema de generadores sea una base. Esto nos proporciona la unicidad de la representación y permite definir las coordenadas de un elemento en una base de forma única.

Sistemas de Generadores y Bases

- Implicitamente un sistema de generadores tiene *redundancia* en la información. La forma de generar un elemento del espacio no es única.
- El cálculo del rango de una matriz es una medida de la información presente en la matriz. Es el número máximo de vectores (columnas) linealmente independientes. El concepto de dimensión de un espacio vectorial como número mínimo de generadores.
- Necesitamos independencia lineal para que un sistema de generadores sea una base. Esto nos proporciona la unicidad de la representación y permite definir las coordenadas de un elemento en una base de forma única.
- Álgebra: Espacios vectoriales de dimensión finita.

Sistemas de Generadores y Bases

- Implicitamente un sistema de generadores tiene *redundancia* en la información. La forma de generar un elemento del espacio no es única.
- El cálculo del rango de una matriz es una medida de la información presente en la matriz. Es el número máximo de vectores (columnas) linealmente independientes. El concepto de dimensión de un espacio vectorial como número mínimo de generadores.
- Necesitamos independencia lineal para que un sistema de generadores sea una base. Esto nos proporciona la unicidad de la representación y permite definir las coordenadas de un elemento en una base de forma única.
- Álgebra: Espacios vectoriales de dimensión finita.
- Cálculo: Espacios vectoriales de dimensión infinita. Espacios funcionales

Bases Ortogonales

- El concepto de ortogonalidad es más fuerte que el concepto de independencia lineal. De hecho la ortogonalidad implica la independencia lineal.

Bases Ortogonales

- El concepto de ortogonalidad es más fuerte que el concepto de independencia lineal. De hecho la ortogonalidad implica la independencia lineal.
- Para calcular los coeficientes (pesos o coordenadas) de un vector en una base es necesario resolver un sistema lo que aumenta la carga computacional del algoritmo.

Bases Ortogonales

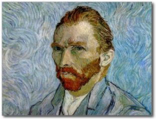
- El concepto de ortogonalidad es más fuerte que el concepto de independencia lineal. De hecho la ortogonalidad implica la independencia lineal.
- Para calcular los coeficientes (pesos o coordenadas) de un vector en una base es necesario resolver un sistema lo que aumenta la carga computacional del algoritmo.
- Para calcular los coeficientes (pesos o coordenadas) de un vector en una base ortogonal es suficiente realizar un producto escalar. Necesitamos la estructura de espacio de Hilbert.

Bases Ortogonales

- El concepto de ortogonalidad es más fuerte que el concepto de independencia lineal. De hecho la ortogonalidad implica la independencia lineal.
- Para calcular los coeficientes (pesos o coordenadas) de un vector en una base es necesario resolver un sistema lo que aumenta la carga computacional del algoritmo.
- Para calcular los coeficientes (pesos o coordenadas) de un vector en una base ortogonal es suficiente realizar un producto escalar. Necesitamos la estructura de espacio de Hilbert.
- Generalización al caso infinito dimensional. Teoría de las series de Fourier. El problema de autofunciones. Desarrollos en series de Fourier mediante autofunciones.

Transformaciones Lineales

- Las Transformaciones Lineales son aplicaciones lineales, $T(\mathbf{x}) = A\mathbf{x}$. Las Transformaciones afines son aplicaciones lineales compuestas con una traslación, $T(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$.



Transformaciones Lineales

- Espacios de arranque y llegada (input-output). Núcleo e Imagen de una aplicación lineal.

Transformaciones Lineales

- Espacios de arranque y llegada (input-output). Núcleo e Imagen de una aplicación lineal.
- La **fórmula de la dimensión**. Sea $A \in \mathbb{R}^{m,n}$ una matriz y $f : \mathbb{R}^n \to \mathbb{R}^m$ la aplicación lineal asociada. Se tiene entonces

$$n = \mathrm{d}im\,\mathcal{R}(A) + \mathrm{d}im\,\mathcal{N}(A)$$

- Espacios Vectoriales Normados. Espacios de Banach e Hilbert. Producto escalar y Normas. Normas y distancias. Energías.

Transformaciones Lineales

- Espacios de arranque y llegada (input-output). Núcleo e Imagen de una aplicación lineal.
- La **fórmula de la dimensión**. Sea $A \in \mathbb{R}^{m,n}$ una matriz y $f : \mathbb{R}^n \to \mathbb{R}^m$ la aplicación lineal asociada. Se tiene entonces

$$n = \mathrm{d}im\,\mathcal{R}(A) + \mathrm{d}im\,\mathcal{N}(A)$$

- Espacios Vectoriales Normados. Espacios de Banach e Hilbert. Producto escalar y Normas. Normas y distancias. Energías.
- Problemas de minimización de la energía. Energías convexas (existencia) y estrictamente convexas (existencia y unicidad).

Contenido

Formas cuadráticas

- Definición, representación y clasificación.

Formas cuadráticas

- Definición, representación y clasificación.
- Clasificación mediante el estudio del signo de la forma cuadrática. Herramientas gráficas.

Formas cuadráticas

- Definición, representación y clasificación.
- Clasificación mediante el estudio del signo de la forma cuadrática. Herramientas gráficas.
- Clasificación mediante el cálculo de los autovalores del problema. La forma canónica de una forma cuadrática.

Formas cuadráticas

- Definición, representación y clasificación.
- Clasificación mediante el estudio del signo de la forma cuadrática. Herramientas gráficas.
- Clasificación mediante el cálculo de los autovalores del problema. La forma canónica de una forma cuadrática.
- Clasificación mediante el estudio del signo de los autovalores del problema.

Formas cuadráticas

- Definición, representación y clasificación.
- Clasificación mediante el estudio del signo de la forma cuadrática. Herramientas gráficas.
- Clasificación mediante el cálculo de los autovalores del problema. La forma canónica de una forma cuadrática.
- Clasificación mediante el estudio del signo de los autovalores del problema.
- El problema de autovalores: $A\mathbf{x} = \lambda\mathbf{x}$. Su resolución implica el estudio y resolución de Sistemas homogéneos.

Formas cuadráticas

- Definición, representación y clasificación.
- Clasificación mediante el estudio del signo de la forma cuadrática. Herramientas gráficas.
- Clasificación mediante el cálculo de los autovalores del problema. La forma canónica de una forma cuadrática.
- Clasificación mediante el estudio del signo de los autovalores del problema.
- El problema de autovalores: $A\mathbf{x} = \lambda\mathbf{x}$. Su resolución implica el estudio y resolución de Sistemas homogéneos.
- Necesitamos una teoría para el diseño de algoritmos que puedan resolver con rapidéz y precisión los sistemas lineales.

El problema de autovalores.

- Se trata de un problema fundamental en álgebra lineal con generalización natural al problema del cálculo de autofunciones en cálculo. Está en la base de las técnicas de descomposición de señales e imágenes.

El problema de autovalores.

- Se trata de un problema fundamental en álgebra lineal con generalización natural al problema del cálculo de autofunciones en cálculo. Está en la base de las técnicas de descomposición de señales e imágenes.
- Permite el estudio de la diagonalización de una matriz y el cálculo de la transformación diagonalizante. Permite desacoplar sistemas de ecuaciones.

El problema de autovalores.

- Se trata de un problema fundamental en álgebra lineal con generalización natural al problema del cálculo de autofunciones en cálculo. Está en la base de las técnicas de descomposición de señales e imágenes.
- Permite el estudio de la diagonalización de una matriz y el cálculo de la transformación diagonalizante. Permite desacoplar sistemas de ecuaciones.
- Permite el cálculo de los operadores de traza y deteminante de una matriz

El problema de autovalores.

- Se trata de un problema fundamental en álgebra lineal con generalización natural al problema del cálculo de autofunciones en cálculo. Está en la base de las técnicas de descomposición de señales e imágenes.
- Permite el estudio de la diagonalización de una matriz y el cálculo de la transformación diagonalizante. Permite desacoplar sistemas de ecuaciones.
- Permite el cálculo de los operadores de traza y deteminante de una matriz
- Permite el estudio de Condicionamiento de un sistema.

El problema de autovalores.

- Se trata de un problema fundamental en álgebra lineal con generalización natural al problema del cálculo de autofunciones en cálculo. Está en la base de las técnicas de descomposición de señales e imágenes.
- Permite el estudio de la diagonalización de una matriz y el cálculo de la transformación diagonalizante. Permite desacoplar sistemas de ecuaciones.
- Permite el cálculo de los operadores de traza y deteminante de una matriz
- Permite el estudio de Condicionamiento de un sistema.
- Permite el cálculo de la Descomposición Espectral. Permite realizar la Descomposición en Valores Singulares (SVD) de una matriz. Aplicación a la compresión digital de imágenes.

Contenido

Sistemas Lineales

- Pueden ser de tipo homogéneo, $A\mathbf{x} = \mathbf{0}$ o no homogéneo $A\mathbf{x} = \mathbf{b}$.

Sistemas Lineales

- Pueden ser de tipo homogéneo, $A\mathbf{x} = \mathbf{0}$ o no homogéneo $A\mathbf{x} = \mathbf{b}$.
- El conjunto de soluciones de un sistema homogéneo es un espacio vectorial.

Sistemas Lineales

- Pueden ser de tipo homogéneo, $A\mathbf{x} = \mathbf{0}$ o no homogéneo $A\mathbf{x} = \mathbf{b}$.
- El conjunto de soluciones de un sistema homogéneo es un espacio vectorial.
- El conjunto de soluciones de un sistema no homogéneo es un espacio afín (estructuras lineales que no pasan por el origen).

Sistemas Lineales

- Pueden ser de tipo homogéneo, $A\mathbf{x} = \mathbf{0}$ o no homogéneo $A\mathbf{x} = \mathbf{b}$.
- El conjunto de soluciones de un sistema homogéneo es un espacio vectorial.
- El conjunto de soluciones de un sistema no homogéneo es un espacio afín (estructuras lineales que no pasan por el origen).
- Resolver eficientemente un problema de autovalores consiste en saber resolver eficientemente un sistema lineal homogéneo.

Sistemas Lineales

- Pueden ser de tipo homogéneo, $A\mathbf{x} = \mathbf{0}$ o no homogéneo $A\mathbf{x} = \mathbf{b}$.
- El conjunto de soluciones de un sistema homogéneo es un espacio vectorial.
- El conjunto de soluciones de un sistema no homogéneo es un espacio afín (estructuras lineales que no pasan por el origen).
- Resolver eficientemente un problema de autovalores consiste en saber resolver eficientemente un sistema lineal homogéneo.
- Dificultad: la solución trivial $\mathbf{x} = \mathbf{0}$ siempre es solución de un sistema homogéneo. Si hay unicidad no nos sirve. Si no hay unicidad hay infinitas soluciones. Necesito imponer restricciones. ejemplo: Calibrado de cámaras.

Sistemas Compatibles e Incompatibles

- **El teorema de Rouché-Frobenius** Permite estudiar el sistema (existencia y unicidad) mediante el estudio del rango de una matriz y su ampliada

Sistemas Compatibles e Incompatibles

- **El teorema de Rouché-Frobenius** Permite estudiar el sistema (existencia y unicidad) mediante el estudio del rango de una matriz y su ampliada
- **Fórmula de la Dimensión** Permite estudiar el sistema como un operador de transformación entre espacios vectoriales. Se basa en el estudio de la imagen y el núcleo de la transformación.

Sistemas Compatibles e Incompatibles

- **El teorema de Rouché-Frobenius** Permite estudiar el sistema (existencia y unicidad) mediante el estudio del rango de una matriz y su ampliada
- **Fórmula de la Dimensión** Permite estudiar el sistema como un operador de transformación entre espacios vectoriales. Se basa en el estudio de la imagen y el núcleo de la transformación.
- Es necesario dar un sentido a los sistemas incompatibles. La mayoría de los sistemas son incompatibles debido al ruidoo inherente al proceso de adquisición de datos. El sentido de los mínimos cuadrados.

Sistemas Indeterminados y Sobredeterminados

- **Sistemas Indeterminados** El número de restricciones (ecuaciones del sistema) es menor que el número de grados de libertad (número de incógnitas del sistema). Aparece el problema de la no unicidad de soluciones. Problemas mal planteados. Admiten infinitas soluciones paramétricas.

Sistemas Indeterminados y Sobredeterminados

- **Sistemas Indeterminados** El número de restricciones (ecuaciones del sistema) es menor que el número de grados de libertad (número de incógnitas del sistema). Aparece el problema de la no unicidad de soluciones. Problemas mal planteados. Admiten infinitas soluciones paramétricas.
- **Sistemas Sobredeterminados** El número de restricciones (ecuaciones del sistema) es mayor que el número de grados de libertad (número de incógnitas del sistema). Aparece el problema de la no existencia de soluciones. Problemas mal planteados. Es necesario dar un sentido *débil* al concepto de solución. La solución de mínimos cuadrados.

Sistemas bien planteados

Los sistema pueden estar **bien planteados** (well-posed) o **mal planteados** (ill-posed).

Los sistemas mal planteados se tienen que regularizar. Numéricamente sólo se pueden resolver problemas bien planteados.

La **regularización de Tikhonov**. Generalización al calculo variacional. Procesado de imágenes. Modelos matemáticos. Formulaciones Bayesianas.

- **Existencia** de soluciones

Sistemas bien planteados

Los sistema pueden estar **bien planteados** (well-posed) o **mal planteados** (ill-posed).

Los sistemas mal planteados se tienen que regularizar. Numéricamente sólo se pueden resolver problemas bien planteados.

La **regularización de Tikhonov**. Generalización al calculo variacional. Procesado de imágenes. Modelos matemáticos. Formulaciones Bayesianas.

- **Existencia** de soluciones
- **Unicidad** de soluciones

Sistemas bien planteados

Los sistema pueden estar **bien planteados** (well-posed) o **mal planteados** (ill-posed).

Los sistemas mal planteados se tienen que regularizar. Numéricamente sólo se pueden resolver problemas bien planteados.

La **regularización de Tikhonov**. Generalización al calculo variacional. Procesado de imágenes. Modelos matemáticos. Formulaciones Bayesianas.

- **Existencia** de soluciones
- **Unicidad** de soluciones
- **El Principio de Hadamard**. Existencia, Unicidad y Dependencia continua entre datos y soluciones de un sistema. Sistemas robustos.

Sistemas bien planteados

Los sistema pueden estar **bien planteados** (well-posed) o **mal planteados** (ill-posed).

Los sistemas mal planteados se tienen que regularizar. Numéricamente sólo se pueden resolver problemas bien planteados.

La **regularización de Tikhonov**. Generalización al calculo variacional. Procesado de imágenes. Modelos matemáticos. Formulaciones Bayesianas.

- **Existencia** de soluciones
- **Unicidad** de soluciones
- **El Principio de Hadamard**. Existencia, Unicidad y Dependencia continua entre datos y soluciones de un sistema. Sistemas robustos.

Contenido

El problema de mínimos cuadrados

Se construye calculando las *ecuaciones normales* asociadas al sistema.

El problema de mínimos cuadrados *siempre* tiene solución. Puede no ser única. El problema puede estar mal planteado.

Su resolución se basa en la teoría de la *pseudo inversa* de Moore-Penrose (que existe siempre).

- Está a la base de las técnicas de **regresión**, lineal y no lineal.

El problema de mínimos cuadrados

Se construye calculando las *ecuaciones normales* asociadas al sistema.

El problema de mínimos cuadrados *siempre* tiene solución. Puede no ser única. El problema puede estar mal planteado.

Su resolución se basa en la teoría de la *pseudo inversa* de Moore-Penrose (que existe siempre).

- Está a la base de las técnicas de **regresión**, lineal y no lineal.

  Permite dar un sentido a las soluciones de un sistema que no admite solución.

Contenido

Métodos Directos vs Métodos Iterativos

- El Álgebra numérica proporciona algoritmos para la resolución de sistemas lineales.

Métodos Directos vs Métodos Iterativos

- El Álgebra numérica proporciona algoritmos para la resolución de sistemas lineales.
- Métodos directos (sistemas pequeños y bien condicionados) y Métodos Iterativos (sistemas grandes y/o mal condicionados).

Métodos Directos vs Métodos Iterativos

- El Álgebra numérica proporciona algoritmos para la resolución de sistemas lineales.
- Métodos directos (sistemas pequeños y bien condicionados) y Métodos Iterativos (sistemas grandes y/o mal condicionados).
- Métodos Iterativos: la necesidad de un precondicionador.

Métodos Directos vs Métodos Iterativos

- El Álgebra numérica proporciona algoritmos para la resolución de sistemas lineales.
- Métodos directos (sistemas pequeños y bien condicionados) y Métodos Iterativos (sistemas grandes y/o mal condicionados).
- Métodos Iterativos: la necesidad de un precondicionador.

Ejemplo

El método de Jacobi, el método de Gauss-Seidel, Relajaciones, el método del gradiente, gradiente conjugado, gradiente precondicionado, PCG, GMRES, etc

e

PARTE II - BLOQUE II: Cálculo y Optimización Multivariable

Emanuele Schiavi, Iván Ramírez, Victoria Ruiz

FUNDAMENTOS MATEMÁTICOS

CURSO 2023-2024

UNIVERSIDAD REY JUAN CARLOS

MÁSTER EN VISIÓN ARTIFICIAL

Contenido

1. Introducción
   - El cálculo y el procesamiento digital de imágenes.
   - Aplicaciones

Contenido

Contenido

Introducción

- El contenido de este II bloque de transparencias es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.

Introducción

- El contenido de este II bloque de transparencias es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.
- El **cálculo** es una rama de las matemáticas que permite operar con las *imágenes digitales*. Cálculo de operadores diferenciales multivariables

Introducción

- El contenido de este II bloque de transparencias es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.
- El **cálculo** es una rama de las matemáticas que permite operar con las *imágenes digitales*. Cálculo de operadores diferenciales multivariables
- Optimización basada en el cálculo del gradiente. Minimización de funciones y funcionales.

Introducción

- El contenido de este II bloque de transparencias es una breve introducción a los conceptos generales que se desarrollarán en esta asignatura.
- El **cálculo** es una rama de las matemáticas que permite operar con las *imágenes digitales*. Cálculo de operadores diferenciales multivariables
- Optimización basada en el cálculo del gradiente. Minimización de funciones y funcionales.
- El Método del Descenso del Gradiente

Introducción

Figura: Derivada parcial I_x de la imagen obtenida con una fórmula de derivación implícita de 7 puntos.

Introducción

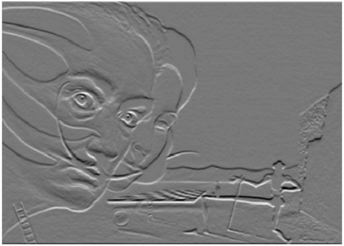


Figura: Derivada parcial I_y de la imagen obtenida con una fórmula de derivación implícita de 7 puntos.

Introducción

Figura: Módulo del Gradiente de la imagen

El cálculo y el procesamiento digital de imágenes.

- El cálculo permite describir no linealidades. La naturaleza es fuertemente no lineal. Cálculo aplicado a funciones (Cálculo Multivariable) y a funcionales (Cálculo Variacional)

El cálculo y el procesamiento digital de imágenes.

- El cálculo permite describir no linealidades. La naturaleza es fuertemente no lineal. Cálculo aplicado a funciones (Cálculo Multivariable) y a funcionales (Cálculo Variacional)
- Permite calcular los operadores de gradiente, divergencia y laplaciano aplicados a una imagen.

Aplicaciones

- Medicina, detección remota, transmisión y codificación de datos, robótica, visión artificial, reconocimiento de patrones, industria cinematográfica, procesamiento de imágenes obtenidas de microscopios y restauración y enfocado de imágenes. Superresolución.
- Detección de bordes y estructuras
- Transformaciones no lineales.

Aplicaciones

- Medicina, detección remota, transmisión y codificación de datos, robótica, visión artificial, reconocimiento de patrones, industria cinematográfica, procesamiento de imágenes obtenidas de microscopios y restauración y enfocado de imágenes. Superresolución.
- Detección de bordes y estructuras
- Transformaciones no lineales.

Contenido

El método del descenso

- Matriz Jacobiana y Matriz Hessiana

El método del descenso

- Matriz Jacobiana y Matriz Hessiana
- El método del descenso revisitado

El método del descenso

- Matriz Jacobiana y Matriz Hessiana
- El método del descenso revisitado
- El método de Newton para minimización

Sistemas no lineales

- Resolución numérica
- El Método de Newton
- Problemas de mínimos cuadrados no lineales
- El Método de descenso gradiente
- El Método de Gauss-Newton
- El Método de Levemberg-Marquard

El Cálculo Variacional

- Definición de un funcional y espacios funcionales (infinito dimensionales). Se plantean modelos continuos de procesamiento de imágenes.

El Cálculo Variacional

- Definición de un funcional y espacios funcionales (infinito dimensionales). Se plantean modelos continuos de procesamiento de imágenes.
- Derivación de un funcional. Derivadas de Frechet y Gateaux
- Condiciones de Optimalidad
- Las ecuaciones de Euler-Lagrange

Cálculo Variacional

- Modelos lineales: Eliminación de ruido. Formulación Bayesiana

Cálculo Variacional

- Modelos lineales: Eliminación de ruido. Formulación Bayesiana
- La regularización de Tikhonov. Generalización al caso no lineal.

Cálculo Variacional

- Modelos lineales: Eliminación de ruido. Formulación Bayesiana
- La regularización de Tikhonov. Generalización al caso no lineal.
- Ecuaciones de Euler-Lagrange

Cálculo Variacional

- Modelos lineales: Eliminación de ruido. Formulación Bayesiana
- La regularización de Tikhonov. Generalización al caso no lineal.
- Ecuaciones de Euler-Lagrange
- Discretización y fórmulas de diferencias finitas

Cálculo Variacional

- Modelos lineales: Eliminación de ruido. Formulación Bayesiana
- La regularización de Tikhonov. Generalización al caso no lineal.
- Ecuaciones de Euler-Lagrange
- Discretización y fórmulas de diferencias finitas
- Modelos no lineales
- Prácticas de procesamiento digital

Introducción al Aprendizaje Automático y Profundo

Iván Ramírez, Victoria Ruiz, Emanuele Schiavi

FUNDAMENTOS MATEMÁTICOS

CURSO 2023-2024

UNIVERSIDAD REY JUAN CARLOS

MÁSTER EN VISIÓN ARTIFICIAL

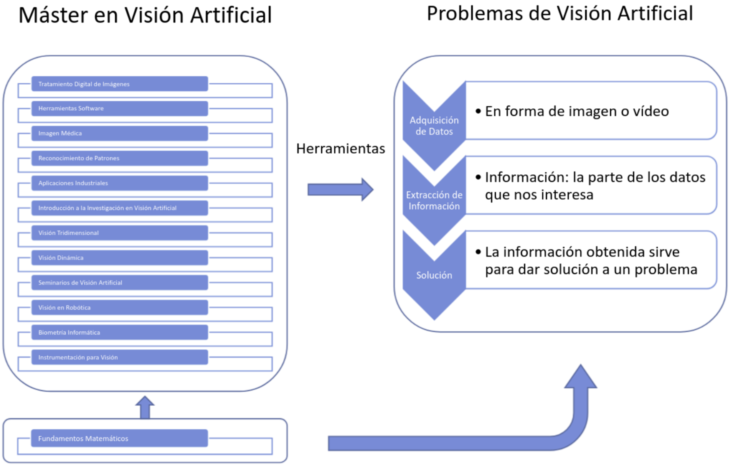
Contenido

Contenido

Contenido

Contenido

Visión General



Modelado: Bayes

Bayes:

$$\underbrace{\mathsf{p}(M/D)}_{\text{Posterior}} = \frac{\overbrace{\mathsf{p}(D/M)}^{\textit{Likelihood}}\,\overbrace{\mathsf{p}(M)}^{Prior}}{\underbrace{\mathsf{p}(D)}_{Evidence}}$$

A partir de aquí se derivan las formulaciones de todos los problemas que se ven, tanto en los Modelos Variacionales de la primera parte de la asignatura, como los problemas de clasificación y regresión en Aprendizaje Automático, así como en Inferencia Bayesiana que veremos en la segunda parte de la misma.

Será clave entender cada uno de los términos que aparecen en la expresión anterior y su significado en cada problema a resolver.

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos
    - Aprendizaje Profundo
    - Problema a resolver: Clasificación y Regresión

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos
  - Aprendizaje Profundo
  - Problema a resolver: Clasificación y Regresión
- Redes Neuronales: función de aproximación universal

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos
  - Aprendizaje Profundo
  - Problema a resolver: Clasificación y Regresión
- Redes Neuronales: función de aproximación universal
- Redes Convolucionales: sesgo inductivo

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos
    - Aprendizaje Profundo
    - Problema a resolver: Clasificación y Regresión
- Redes Neuronales: función de aproximación universal
- Redes Convolucionales: sesgo inductivo
- Descenso de Gradiente: Backpropagation Algorithm

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos
    - Aprendizaje Profundo
    - Problema a resolver: Clasificación y Regresión
- Redes Neuronales: función de aproximación universal
- Redes Convolucionales: sesgo inductivo
- Descenso de Gradiente: Backpropagation Algorithm
- Optimizadores (de primer orden)

Contenidos

Contenidos:

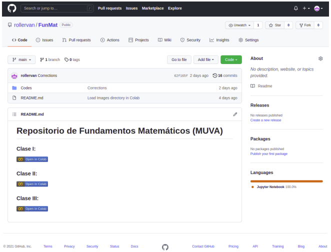
- Introducción informal del Aprendizaje Automático y de sus elementos
    - Aprendizaje Profundo
    - Problema a resolver: Clasificación y Regresión
- Redes Neuronales: función de aproximación universal
- Redes Convolucionales: sesgo inductivo
- Descenso de Gradiente: Backpropagation Algorithm
- Optimizadores (de primer orden)
- Inferencia Bayesiana: Inferencia Variacional (dropout)

Contenidos

Contenidos:

- Introducción informal del Aprendizaje Automático y de sus elementos
  - Aprendizaje Profundo
  - Problema a resolver: Clasificación y Regresión
- Redes Neuronales: función de aproximación universal
- Redes Convolucionales: sesgo inductivo
- Descenso de Gradiente: Backpropagation Algorithm
- Optimizadores (de primer orden)
- Inferencia Bayesiana: Inferencia Variacional (dropout)
- Prácticas

Como alternativa a Matlab, está disponible en FunMat - GitHub un repositorio con los ejercicios y códigos de la asignatura r0.4



- Mismo contenido que en Matlab
- Uso de múltiples librerías:
    - numpy
    - sympy
    - matplotlib ...
- Colab (entorno preinstalado, GPU, etc)

Recomendaciones:

- Entendemos que el uso de Matlab puede, en general, ser más simple. No obstante, si estás más familiarizado/a con Python y no tienes dificultades en usarlo, esta puede ser una buena opción.
- El objetivo de la asignatura no es mejorar o adquirir habilidades de programación, por lo que, si el uso de Python te supone un esfuerzo extra, Matlab debería ser mejor opción.
- No obstante, apoyarse en códigos es muy útil para entender en profundidad los conceptos de la asignatura, por lo que os animamos encarecidamente a que le dediquéis tiempo y practiquéis lo máximo posible.

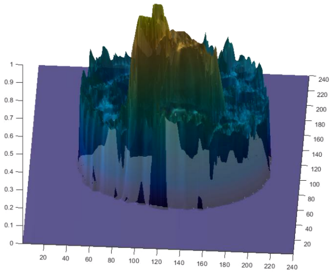
Para quién está pensado:

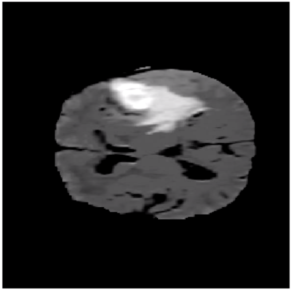
- Perfil informático $\rightarrow$Python
- Perfil no-informático $\rightarrow$Matlab

Ejemplos: Sustracción de Fondo

Ejemplos: Segmentación de Glioblastomas

[width=9.75cm,height=4cm,





poster]

seg$_tum$

Ejemplos: Restauración

Ejemplos: Estimación de Pose 3D

Ejemplos: Self Super-Resolution