

vuela

andaluciavuela.es

CONECTA CON ANDALUCÍA

**PLATAFORMA
DIGITAL DE
ANDALUCÍA**



Junta de Andalucía

MATEMÁTICAS BÁSICAS

ÍNDICE DE CONTENIDO

1.	Introducción y objetivos	3
2.	Introducción a la Teoría de Conjuntos	4
2.1.	Operaciones con conjuntos	5
2.2.	Propiedades de los conjuntos	9
2.3.	Aplicaciones entre conjuntos	10
3.	Análisis matemático	13
3.1.	Concepto de función	13
3.2.	Concepto de límite	15
3.3.	Concepto de derivada	17
3.4.	Concepto de integral	18
4.	Álgebra matricial	19
4.1.	Concepto de matriz	19
4.2.	Operaciones con matrices	21
5.	Estadística descriptiva	25
5.1.	Medidas de posición	29
5.2.	Medidas de dispersión	36
5.3.	Análisis de datos exploratorio	38
5.4.	Medidas estadísticas entre varias variables	41
6.	Conceptos básicos de probabilidad	45
7.1.	Distribuciones de probabilidad	47
7.2.	Tests de hipótesis estadísticos	49
7.	Métodos de Monte-Carlo	51
8.	Nociones de estadística para clasificación binaria	54

1. Introducción y objetivos

En este documento se presentan algunas nociones de conceptos matemáticos que se usarán de forma reiterada en el curso. En particular:

- | Conceptos básicos de Análisis Matemático: Repaso de conjuntos, funciones, límites, derivadas, integrales, etc.
- | Conceptos básicos de Álgebra lineal: En particular, operaciones con matrices.
- | Conceptos básicos de Estadística, diferenciando entre elementos de estadística descriptiva y teoría de la probabilidad.

El objetivo de este documento persigue dos cuestiones fundamentales:

- | Refrescar los conocimientos previos requeridos sobre matemáticas, de modo que se puedan comprender mejor los contenidos del resto del curso.
- | Introducir la notación matemática a utilizar en los demás módulos.

2. Introducción a la Teoría de Conjuntos

Un **conjunto** es una colección de elementos que guardan una relación entre sí. Dichos elementos pueden tener una naturaleza que depende de aquello que se quiera representar, como por ejemplo: El conjunto de las personas que viven en una comunidad de vecinos, el conjunto de colores usado en señales de tráfico, el conjunto de números de 0 al 10, etc. Un **subconjunto** está formado por algunos elementos de otro conjunto original.

Los conjuntos pueden definirse de forma **explícita (por extensión)** o de forma **implícita (por comprensión)**. Se dice que el conjunto se define de forma **explícita** cuando se enumeran cada uno de los elementos del conjunto, mientras que se define de forma **implícita** cuando se remarca alguna propiedad, o algún conjunto de propiedades, que comparten todos los elementos del conjunto. Por ejemplo:

Ejemplo: Definición por extensión del conjunto A, formado por los números enteros del 0 al 10

$$A = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Ejemplo: Definición por comprensión del conjunto B, formado por las vocales de la palabra “manzana”

$$B = \{x : x \text{ es vocal de la palabra manzana}\}$$

(Su definición explícita sería: $B = \{a\}$)

Existen diferentes tipos de conjuntos de acuerdo a diferentes categorías: **Numerables** (se puede asignar un número natural a cada elemento del conjunto) o **no numerables, finitos** (se podría nombrar cada elemento del mismo) o **infinitos, ordenados** (los elementos del conjunto mantienen una relación de orden entre ellos) o **no ordenados**, etc. Un conjunto particular es el denominado **conjunto vacío**, representado por el símbolo \emptyset . El conjunto vacío no tiene ningún elemento.

Ejemplo: Sea C el conjunto de los alumnos matriculados en este curso, entonces...

- ¿Es el conjunto numerable, o no numerable?
- ¿Es el conjunto finito, o infinito?
- ¿Es el conjunto ordenado, o no ordenado?

Conjuntos de números

Los conjuntos numéricos son una parte central en las matemáticas aplicadas a la ingeniería. De especial interés para el curso serán los siguientes conjuntos:

- | **Números naturales:** Notados como $\mathbb{N} = \{0, 1, 2, 3, 4, 5, \dots\}$
- | **Números enteros:** Notados como $\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \dots\}$
- | **Números racionales:** Notados como $\mathbb{Q} = \{x/y : x, y \in \mathbb{Z}, y \neq 0\}$
- | **Números irracionales:** Notados como I . Comprende aquellos números que no se pueden expresar como naturales, enteros o racionales (por ejemplo, los números π, e, \dots).
- | **Números reales:** Notados como $\mathbb{R} = \mathbb{Q} \cup I$
- | **Números complejos:** Notados como $\mathbb{C} = \{a + b\sqrt{-1} : a, b \in \mathbb{R}\}$

2.1. Operaciones con conjuntos

Se pueden definir diversas operaciones entre conjuntos, y también entre elementos particulares y conjuntos. Las operaciones entre conjuntos, a su vez, pueden ser **unarias** (involucran sólo a un conjunto) o **binarias** (involucran a dos conjuntos). Con esta clasificación, las operaciones entre conjuntos son las siguientes:

Complemento de un conjunto A: Operación unaria, notada como \bar{A} , indica todos los demás elementos existentes en el mundo (o problema a resolver) que no pertenecen al conjunto A.

Ejemplo: Complemento del conjunto A de vocales que aparecen en la palabra “manzana”

$$A = \{x : x \text{ es vocal de la palabra manzana}\} = \{a\}$$

$$\bar{A} = A^c = \{e, i, o, u\}$$

Unión de dos conjuntos A y B: Notado como $A \cup B$, da como resultado un nuevo conjunto formado por todos los elementos de los conjuntos A y B (**sin repetir**).

Ejemplo: Unión de los conjuntos A (vocales de la palabra “manzana”) y B (vocales de la palabra “murciélago”)

$$A = \{a\}$$

$$B = \{a, e, i, o, u\}$$

$$A \cup B = \{a, e, i, o, u\}$$

Intersección de dos conjuntos A y B: Notado como $A \cap B$, da como resultado un nuevo conjunto formado por todos los elementos que son comunes o que aparecen en ambos conjuntos A y B (**sin repetir**).

Ejemplo: Intersección de los conjuntos A (vocales de la palabra “manzana”) y B (vocales de la palabra “pera”)

$$A = \{a\}$$

$$B = \{a, e\}$$

$$A \cap B = \{a\}$$

Diferencia entre conjuntos A y B: Notado como $A \setminus B$, da como resultado todos los elementos de A, excluyendo a aquellos elementos de A que también estén en B. **OJO: La operación no tiene la propiedad conmutativa**, por lo que el conjunto $A \setminus B$ puede ser distinto de $B \setminus A$.

Ejemplo: Diferencia de los conjuntos A (vocales de la palabra “mecachis”) y B (vocales de la palabra “vaso”)

$$\begin{aligned}A &= \{a, e, i\} \\ B &= \{a, o\} \\ A \setminus B &= \{e, i\} \\ B \setminus A &= \{o\}\end{aligned}$$

Partes del conjunto A: Operación unaria, notada como $P(A)$. Da como resultado un conjunto, formado por todos los posibles subconjuntos (incluyendo el vacío) de las combinaciones posibles de los elementos de A.

Ejemplo: Partes del conjunto A (vocales de la palabra “mecachis”), incluyendo al conjunto vacío

$$\begin{aligned}A &= \{a, e, i\} \\ P(A) &= \{\emptyset, \{a\}, \{e\}, \{i\}, \{a, e\}, \{a, i\}, \{e, i\}, \{a, e, i\}\}\end{aligned}$$

Cardinal del conjunto A: Operación unaria, notada como $|A|$. Da como resultado el número de elementos contenidos en el conjunto (o infinito, en su caso).

Ejemplo: Cardinal de A (vocales de la palabra “mecachis”), incluyendo al conjunto vacío

$$\begin{aligned}A &= \{a, e, i\} \\ |A| &= 3\end{aligned}$$

Producto cartesiano entre dos conjuntos A y B: Operación binaria, notada como $A \times B$. Da como resultado un conjunto, formado por todas las parejas posibles entre los elementos de A y los elementos de B.

Ejemplo: Producto cartesiano entre el conjunto $A=\{1, 2, 3\}$ y el conjunto $B=\{a, b\}$

$$A \times B = \{(1, a), (1, b), (2, a), (2, b), (3, a), (3, b)\}$$

Aparte de estas operaciones básicas entre conjuntos, también existen **operadores relacionales**. Un operador relacional permite conocer si la relación entre dos conjuntos existe (valor verdadero) o no (valor falso). Entre los operadores relacionales más comunes, tenemos la **equivalencia**, la **inclusión** y sus opuestas:

Equivalencia y desigualdad entre dos conjuntos A y B: Notadas como $A \equiv B$ y $A \neq B$, respectivamente. La relación de igualdad es cierta si tanto A como B contienen exactamente los mismos elementos tanto en un conjunto como en otro. La relación de desigualdad es la opuesta

Ejemplo: Equivalencia entre los conjuntos $A=\{\text{vocales de la palabra "pera"}\}$, $B=\{\text{vocales de la palabra "madre"}\}$, y $C=\{\text{vocales de la palabra "mero"}\}$

$$\begin{aligned} A &= \{a, e\}; B = \{e, a\}; C = \{e, o\} \\ A &\equiv B \text{ es cierto, } A \neq B \text{ es falso} \\ A &\equiv C \text{ es falso, } A \neq C \text{ es cierto} \end{aligned}$$

Inclusión de un conjunto A en un conjunto B: Notada como $A \subset B / A \subseteq B$. La relación $A \subset B$ indica que todos los elementos de A están incluidos en B, y que además B contiene al menos algún elemento más (no contenido en A). Por otra parte, $A \subseteq B$ será cierta si $A \subset B$ o si $A \equiv B$. Los correspondientes operadores opuestos son $A \not\subset B$ (A no está incluido en B), y $A \not\subseteq B$ (A no está incluido en B, y A no es igual a B).

Ejemplo: Relación de inclusión

$$\begin{aligned}A &= \{a, e\}; B = \{e, a\}; C = \{e, o\} \\(B \cap C) &\subset A \text{ es cierto} \\A &\subseteq B \text{ es cierto} \\(B \cup C) &\not\subset A \text{ es cierto}\end{aligned}$$

Finalmente, los **operadores entre elementos y conjuntos** se resumen en el operador de pertenencia y su opuesto. También se consideran operadores relacionales, dado que su comprobación puede dar como resultado valores cierto o falso:

Inclusión de un elemento a en un conjunto A : Notado como $a \in A$. La relación $a \in A$ es cierta si el elemento a pertenece al conjunto A , y falsa en caso contrario. Su correspondiente operador opuesto se nota como $a \notin A$ (el elemento a no pertenece a A).

Ejemplo: Relación de inclusión de elemento en conjunto

$$\begin{aligned}A &= \{a, e\}; B = \{e, a\}; C = \{e, o\} \\e &\in (B \cap C) \text{ es cierto} \\a &\notin B \text{ es falso}\end{aligned}$$

2.2. Propiedades de los conjuntos

En esta sección revisaremos algunas propiedades interesantes de los conjuntos. Asumiremos un conjunto cualquiera X . Entonces, para cada subconjunto A, B, C de X , se cumple:

| **Propiedad conmutativa:** $A \cap B = B \cap A$; $A \cup B = B \cup A$

| **Propiedad asociativa:** $(A \cap B) \cap C = A \cap (B \cap C)$, y también $(A \cup B) \cup C = A \cup (B \cup C)$

| **Propiedad de idempotencia:** $A \cup A = A$; $A \cap A = A$

| **Propiedad de absorción:** $A \cup (A \cap B) = A$; $A \cap (A \cup B) = A$

| **Propiedad Distributiva:** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$, y también $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

| **Leyes de Morgan:** $\overline{(A \cup B)} = \bar{A} \cap \bar{B}$; $\overline{(A \cap B)} = \bar{A} \cup \bar{B}$

2.3. Aplicaciones entre conjuntos

Dados dos conjuntos **X** e **Y**, una aplicación **f** desde el conjunto **X** al conjunto **Y** representa una relación entre ambos conjuntos, que se expresa a su vez como un subconjunto del producto cartesiano entre **X** e **Y**, o $f \subseteq X \times Y$. Una aplicación **f** se suele notar como $f: X \rightarrow Y$. Al conjunto **X** se le denomina **dominio**, mientras que al conjunto **Y** se le conoce como **rango o codominio** de la aplicación.

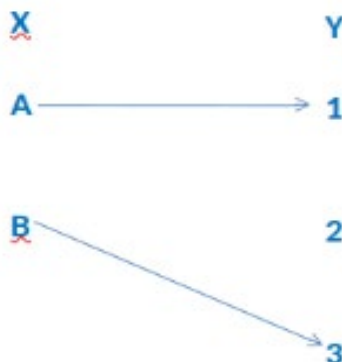
Las aplicaciones sirven para enlazar, o relacionar, elementos del dominio $x \in X$ con otro elemento del rango $y \in Y$. Al elemento que le corresponde al elemento **x** en **Y** mediante la aplicación **f** se le nota como **y=f(x)**. Al subconjunto de elementos de **Y** tal que hay un $x \in X, y \in Y: f(x) = y$ se le conoce como **imagen de f**, y se nota como **Img(f)**.

Ejemplo: Relación entre los colores de un semáforo y su posición

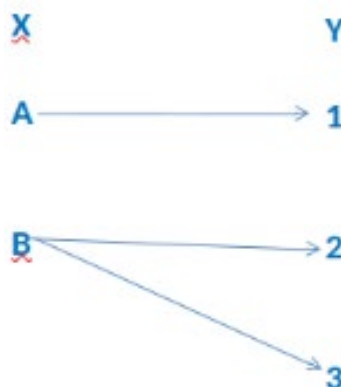
$$\begin{aligned}
 C &= \{\text{rojo, verde, amarillo}\} \\
 P &= \{\text{arriba, centro, abajo}\} \\
 f &: C \rightarrow P \\
 f &= \{(\text{rojo, arriba}), (\text{verde, abajo}), \\
 f(\text{rojo}) &= \text{arriba}; f(\text{verde}) = \text{abajo}; f(\text{amarillo}) = \text{centro}
 \end{aligned}$$

Algunas propiedades de las aplicaciones que serán útiles a continuación son;

| **Una aplicación f es inyectiva** si, dados dos elementos $x_1, x_2 \in X: x_1 \neq x_2$, entonces $f(x_1) \neq f(x_2) \forall x_1, x_2$. Es decir, si no hay ningún elemento de Y que tenga asociado más de un elemento de X .



| **Una aplicación f es sobreyectiva** si $Im g(f) = Y$; es decir, si no existe ningún elemento de Y que no tenga su correspondiente en X .



| **Una aplicación f es biyectiva** si es inyectiva y sobreyectiva.

La **Composición de aplicaciones** consiste en llevar a cabo varias aplicaciones de un conjunto a otro en secuencia. Dadas dos aplicaciones $f: X \rightarrow Y, g: Y \rightarrow Z$, la composición de f con g se nota como $g \circ f$ y significa $g(f(x))$, donde $x \in X$.

Ejemplo: Composición de aplicaciones entre la posición de un color en un semáforo y su significado

$$\begin{aligned}C &= \{\text{rojo, verde, amarillo}\} \\P &= \{\text{arriba, centro, abajo}\} \\S &= \{\text{pasar, parar, cuidado}\} \\f: P &\rightarrow C, g: C \rightarrow S \\f &= \{(\text{arriba, rojo}), (\text{abajo, verde}), (\text{centro, amarillo})\} \\g &= \{(\text{rojo, parar}), (\text{verde, pasar}), (\text{amarillo, cuidado})\} \\(g \circ f)(\text{arriba}) &= \text{parar}; (g \circ f)(\text{abajo}) = \text{pasar} \\(g \circ f)(\text{centro}) &= \text{cuidado}\end{aligned}$$

3. Análisis matemático

En este apartado se realiza un repaso de los conceptos básicos sobre funciones, límites, derivadas, integrales, etc., de modo que se asienten los conocimientos necesarios en módulos posteriores de aprendizaje automático.

3.1. Concepto de función

Una **función** matemática es **una aplicación entre conjuntos**, donde estos deben ser forzosamente **conjuntos numéricos**. Usualmente, las funciones vienen determinadas por una **expresión** algebraica, que involucra distintos operadores como suma, producto, exponenciación, logaritmos, cambio de signo, etc.

Una **función f de una variable real** es una aplicación donde el dominio es el conjunto R . Usualmente, en los problemas que trataremos el rango también será R . Las funciones nos servirán para establecer relaciones entre datos de los problemas que tendremos que resolver, principalmente entre datos de entrada (dominio) y datos de salida (imagen).

Ejemplo: Definición de una función de una variable real mediante su expresión

$$\begin{aligned} f: R &\rightarrow R \\ f(x) &= \log_2(x^2) + 7x \end{aligned}$$

Igualmente importante es el concepto de **función inversa**. La inversa de una función **f** se nota como f^{-1} . Su dominio es el rango de la función **f** , y su codominio es el dominio de la función **f** :

$$\begin{aligned} f: X &\rightarrow Y \\ f^{-1}: Y &\rightarrow X \end{aligned}$$

Si la función **f** asigna un elemento $y \in Y$ a un elemento del conjunto $x \in X$ mediante la aplicación $y=f(x)$, entonces la función inversa asigna un elemento de la imagen $Img(f)$ al conjunto **X**, de modo que $x = f^{-1}(y) \leftrightarrow y = f(x)$. Cabe destacar que **no todas las funciones son invertibles**.

Ejemplo: Función invertible

$$\begin{aligned}f: R^+ &\rightarrow R \\ f(x) &= 2^x \\ f^{-1}(x) &= \log_2(x)\end{aligned}$$

A lo largo de todo el curso, trabajaremos principalmente con funciones de más de una variable. Supondremos, en un caso genérico, que tenemos un total de **n** variables de entrada (dominio) y una variable de salida. Este tipo de funciones de más de una variable real se modelan como sigue:

$$f: R^n \rightarrow R$$

Donde R^n significa *el producto cartesiano de R* un total de **n** veces. El caso más simple son las funciones de dos variables reales:

$$f: R^2 \rightarrow R = f: R \times R \rightarrow R$$

Ejemplo: Definición de una función de dos variables reales mediante su expresión algebraica

$$\begin{aligned}f: R^2 &\rightarrow R \\ f(x, y) &= x + y\end{aligned}$$

Cabe destacar que, por lo general, las funciones reales de varias variables no suelen ser invertibles.

3.2. Concepto de límite

El concepto de **límite en un punto** de una función real de variable real se relaciona con cómo varían los valores de la función a medida que el valor de la variable se acerca a un determinado punto. El valor del límite puede pertenecer al rango de la función, aunque también puede no pertenecer a él.

Notaremos que el límite de la función **f** es el valor **b**, cuando su variable **x tiende** a un valor **a**, de la siguiente forma:

$$\lim_{x \rightarrow a} f(x) \rightarrow b$$

El límite de una función real de variable real en un punto puede existir, y por tanto ser un valor real, o puede no existir y tomar valores como $+\infty$ o $-\infty$.

Ejemplo: Límite de la función logaritmo cuando x tiende a 0

$$\begin{aligned} f: R &\rightarrow R \\ f(x) &= \log(x) \\ \lim_{x \rightarrow 0} f(x) &\rightarrow -\infty \end{aligned}$$

Ejemplo: Límite de la función sigmoide cuando x tiende a infinito

$$\begin{aligned} f: R &\rightarrow R \\ f(x) &= \frac{1}{1 + e^x} \\ \lim_{x \rightarrow \infty} f(x) &\rightarrow 0 \end{aligned}$$

Existen algunas reglas que facilitan el cálculo de límites, tales como:

El límite de la suma (o resta) de dos funciones:

$$\lim_{x \rightarrow a} (f + g)(x) = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x)$$

El límite del producto de dos funciones:

$$\lim_{x \rightarrow a} (f \cdot g)(x) = \lim_{x \rightarrow a} f(x) \cdot \lim_{x \rightarrow a} g(x)$$

El límite del cociente de dos funciones:

$$\lim_{x \rightarrow a} (f/g)(x) = \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)}$$

El límite de la potencia:

$$\lim_{x \rightarrow a} f(x)^{g(x)} = \lim_{x \rightarrow a} f(x)^{\lim_{x \rightarrow a} g(x)}$$

Existen también casos particulares, denominados **indeterminaciones**, que se dan cuando:

$$\lim_{x \rightarrow a} f(x) = \frac{0}{0}$$

$$\lim_{x \rightarrow a} f(x) = \frac{\pm\infty}{\pm\infty}$$

$$\lim_{x \rightarrow a} f(x) = 0 \cdot \pm\infty$$

$$\lim_{x \rightarrow a} f(x) = \infty - \infty$$

Como es bien conocido, existen mecanismos y procedimientos que permiten tratar cada una de estas indeterminaciones para poder calcular el valor del límite de una función. No obstante, estos procedimientos quedan fuera del ámbito de este curso.

3.3. Concepto de derivada

El concepto intuitivo de derivada de una función real de variable real se asocia a tratar de conocer cómo cambia o varía esa función a lo largo de su dominio. Por ejemplo, conocer cómo varía la posición de un vehículo con respecto al tiempo nos proporciona el concepto de velocidad: La derivada del espacio con respecto al tiempo. Si queremos conocer cómo varía dicha velocidad con respecto al tiempo, obtenemos el concepto de aceleración (derivada de la velocidad con respecto al tiempo). La idea de derivada y su cálculo será importante y una parte central en muchos de los métodos utilizados en el aprendizaje automático dentro de la inteligencia artificial; de hecho, es el principal mecanismo utilizado para elaborar algoritmos de aprendizaje de redes neuronales artificiales y técnicas de Deep Learning.

Formalmente, la **derivada de una función real de variable real, continua, en un punto a** se define a partir del siguiente límite:

$$f'(a) = \frac{\partial f}{\partial x}(a) = \lim_{x \rightarrow a} \frac{f(a) - f(x)}{a - x}$$

Si el límite anterior no existe, entonces **la función no es derivable en el punto a** .

Existen varias reglas de derivación, ampliamente conocidas, que nos permiten simplificar el cálculo del límite anterior para obtener la expresión algebraica de la derivada de una función. Por ejemplo, para una función lineal de tipo $f(x) = ax + b$, la regla de derivación es $f'(x) = a$. Otro ejemplo se da para funciones con variable en potencias, como $f(x) = x^n, n \in \mathbb{N}$, en cuyo caso la derivada se calcula como $f'(x) = nx^{n-1}$. Tanto las técnicas de derivación como las reglas del cálculo de derivadas quedan fuera del contexto de este curso, aunque el concepto de derivada sí será necesario en módulos posteriores.

3.4. Concepto de integral

Una integral puede verse conceptualmente como la operación inversa a la derivada de una función continua real de variable real. Existen dos conceptos en integrales básicos: **primitiva** e **integral indefinida**.

Sean dos funciones reales continuas de variable real, definidas como $F: R \rightarrow R, f: R \rightarrow R$. Si **f** es la función derivada de **F**, entonces se dice que **F es la primitiva de f**; es decir, la función que, cuando la derivamos, da **f** como resultado. La **integral indefinida** de **f** con respecto a su variable real **x** se nota como:

$$\int f(x)dx = F(x) + c$$

Donde $c \in R$ es un valor constante numérico desconocido a priori.

Ejemplo: Primitiva de una función

$$f: R \rightarrow R$$

$$f(x) = 3$$

$$\int f(x)dx = 3x + c$$

Las funciones $F(x)=3x+2$ o $G(x)=3x+27$ son ejemplos de primitivas de $f(x)$

Desde el punto de vista geométrico, una integral también puede verse como el área existente entre los valores de una función y el eje de coordenadas sobre el que se muestra. Este concepto será de utilidad en futuros módulos para poder definir conceptos tales como curvas ROC o AUC. En este punto, podemos definir el término de **integral definida** entre dos puntos a y b como:

$$\int_a^b f(x)dx = F(b) - F(a)$$

Existen varias técnicas/metodologías, y también reglas básicas para realizar el cálculo de integrales. Es un área extensa, al igual que la derivación y el cálculo diferencial, que queda fuera del ámbito de este curso. Para nosotros únicamente será necesario conocer el concepto intuitivo de integral, y usarlo debidamente en los análisis de datos que debemos hacer dentro de la solución de problemas con aprendizaje automático e inteligencia artificial.

4. Álgebra matricial

Una de las herramientas más potentes dentro de la Inteligencia Artificial y el aprendizaje automático se basa en el álgebra computacional. El álgebra computación es un área de conocimiento que involucra conceptos propios del álgebra (espacios vectoriales, matrices, cuerpos finitos, etc.), junto con algoritmos de cálculo que operan sobre este tipo de construcciones matemáticas. En este apartado, nos centraremos principalmente en el concepto de matriz y operaciones con matrices, dado que serán las principales herramientas a usar en los siguientes módulos.

4.1. Concepto de matriz

Una matriz es un grupo de números organizados en filas y columnas. Habitualmente, se suele identificar como una rejilla de **n** filas y **m** columnas delimitadas por paréntesis:

$$A_{n,m} = \begin{pmatrix} a_{1,1} & \dots & a_{1,m} \\ \dots & \dots & \dots \\ a_{n,1} & \dots & a_{n,m} \end{pmatrix}$$

El término $A_{n,m}$ indica que **A** es una matriz, con **n** filas y **m** columnas. Cada celda de una matriz en la fila **i** y la columna **j** contiene un valor, que escribimos como $a_{i,j}$. La **dimensión** de una matriz viene determinada por su número de filas y de columnas.

Ejemplo: Matriz de números reales de dimensión 2x3

$$A = \begin{pmatrix} 1.1 & 2.4 & 3.2 \\ 3.7 & 8.4 & 4.5 \end{pmatrix}$$

Algunas matrices, por sus características, tienen nomenclatura específica:

- | **Matriz cuadrada:** Matriz que tiene el mismo número de filas que de columnas.
- | **Matriz diagonal:** Todas las celdas tienen valores 0, salvo algún o algunos elementos en la diagonal (celdas colocadas en posiciones $a_{i,i}$).
- | **Matriz identidad:** Es una matriz cuadrada diagonal donde todos los elementos de la diagonal tienen valor 1. Se nota como I_n (**n** filas y **n** columnas).

Ejemplo: Matriz identidad I_3

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

- | **Matriz traspuesta:** Si **A** es una matriz, su matriz traspuesta se nota como **A^t**, y se calcula intercambiando filas por columnas.

Ejemplo: Matriz traspuesta

$$A = \begin{pmatrix} 1.1 & 2.4 & 3.2 \\ 3.7 & 8.4 & 4.5 \end{pmatrix}$$
$$A^t = \begin{pmatrix} 1.1 & 3.7 \\ 2.4 & 8.4 \\ 3.2 & 4.5 \end{pmatrix}$$

Matriz adjunta: Si A es una matriz, su matriz adjunta se nota como A^\dagger , y se calcula intercambiando filas por columnas, y cambiando cada elemento por su conjugado. La matriz adjunta es de especial interés cuando se trabaja con números complejos (y básica en computación cuántica). Se recuerda que el conjugado de un número complejo $a+bi$ se calcula como $a-bi$. En el caso de números reales, la matriz adjunta es también la matriz traspuesta.

Ejemplo: Matriz adjunta

$$A = \begin{pmatrix} 1+i & 2-4i & 3+2i \\ 3+7i & 8+4i & 4 \\ 1-1 & 3-7i & 4 \end{pmatrix}$$
$$A^\dagger = \begin{pmatrix} 1-1 & 2+4i & 3-2i \\ 3-7i & 8-4i & 4 \\ 4 & 4 & 4 \end{pmatrix}$$

Matriz inversa: Si A es una matriz cuadrada, su matriz inversa se nota como A^{-1} , y es aquella matriz tal que se cumple: $AA^{-1} = A^{-1}A = I$; es decir, el producto de una matriz por su inversa da como resultado la matriz identidad (de la misma dimensión).

4.2. Operaciones con matrices

En este apartado revisaremos las operaciones más comunes con matrices, así como el cálculo de algunas propiedades de las mismas.

Suma y resta de matrices

Sean dos matrices $A_{n,m}, B_{n,m}$, de la misma dimensión $n \times m$. La suma/resta de ambas matrices da como resultado otra matriz de la misma dimensión, donde cada una de las celdas $c_{i,j}$ se calcula como:

$$c_{i,j} = a_{i,j} + b_{i,j} \text{ (suma)}$$

$$c_{i,j} = a_{i,j} - b_{i,j} \text{ (resta)}$$

Ejemplo: Suma de matrices

$$A = \begin{pmatrix} 1+i & 2-4i & 3+2i \\ 3+7i & 8+4i & 4 \end{pmatrix}$$

$$B = \begin{pmatrix} -i & 4i & -2i \\ -7i & -4i & 0 \end{pmatrix}$$

$$C = A + B = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 8 & 4 \end{pmatrix}$$

Multiplicación de matrices

Sean dos matrices $A_{n,k}$, $B_{k,m}$ (el número de columnas de **A** debe obligatoriamente ser el número de filas de **B**). Entonces, se define el producto/multiplicación de matrices **A*****B** y da como resultado una matriz **C**, de dimensión $C_{n,m}$ (número de filas de **A** y número de columnas de **B**), donde cada celda $c_{i,j}$ de la matriz resultante se calcula como:

$$c_{i,j} = \sum_{p=1}^k a_{i,p} * b_{p,j}$$

Es común recordar la multiplicación de matrices mediante la regla conocida como "*filas por columnas*", dado que cada elemento $c_{i,j}$ se calcula como multiplicando uno a uno cada elemento de la fila **i** de la matriz **A** por el elemento de la columna **j** de la matriz **B** y sumando todos los resultados obtenidos.

Ejemplo: Multiplicación de matrices

$$\begin{aligned}
 A_{2,3} &= \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\
 B_{3,2} &= \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \\
 C_{2,2} = A * B &= \begin{pmatrix} 14 & 32 \\ 32 & 77 \end{pmatrix}
 \end{aligned}$$

Determinante de una matriz

El determinante de una matriz cuadrada sirve como medida de independencia de las filas/columnas de la matriz. En nuestro caso, servirá para identificar si existen variables o datos que puedan ser descartados para los cálculos en métodos de aprendizaje automático, dado que una matriz, de tener determinante igual a 0, indicará que hay filas (datos) o columnas (variables) que pueden expresarse como combinación lineal de otras.

El determinante de una matriz cuadrada **A** de números reales, de dimensión **n x n**, se nota como **|A|**, y también es un valor real. Formalmente, se calcula como:

$$|A| = \sum_{j=1}^n a_{i,j} (-1)^{i+j} |A_{i,j}|$$

Donde $A_{i,j}$ significa la submatriz de **A** resultante de eliminar la fila **i** y la columna **j**.

Para matrices de tamaños pequeños, existen reglas automatizadas que calculan el determinante de forma rápida y sin necesidad de aplicar métodos recursivos, como por ejemplo para matrices de dimensión **2x2**, **3x3** o **4x4**. El estudio de estas reglas queda fuera del contexto de este curso.

Ejemplo: Cálculo de determinantes

$$A_{2,2} = \begin{pmatrix} 1 & 2 \\ -2 & 3 \end{pmatrix}$$
$$|A| = 1 * 3 - 2 * (-2) = 7$$

Rango de una matriz

El rango de una matriz es una medida un poco más fina que el determinante, para indicar si las filas/columnas de una matriz pueden expresarse como combinación lineal de otras. Formalmente, el rango de una matriz se define como el orden máximo de los menores de la matriz que no son 0. Equivale al número máximo de filas/columnas de la matriz que son independientes entre sí.

No existe un método particular que sea eficiente en todos los casos para el cálculo del rango de una matriz. Por lo general, depende de si la matriz contiene variables desconocidas (en este caso una descomposición Gauss-Jordan para transformar la matriz en forma escalonada puede conducir rápidamente al cálculo del rango), de si no tiene variables pero es matriz cuadrada (en este caso el cálculo de determinantes de submatrices puede ser un mecanismo fácil), o de si no tiene variables ni es cuadrada (método del cálculo de menores).

Para el curso actual, utilizaremos software numérico que nos permita calcular directamente el rango de matrices, sin entrar en detalles de cómo se realizan estos cálculos internamente. Sí será de interés, no obstante, conocer las implicaciones previamente comentadas entre el valor del rango de una matriz y su relación con la independencia de filas/columnas.

Cálculo de la inversa de una matriz

Dada una matriz cuadrada **A**, cuyo **determinante es distinto de 0**, existe una matriz A^{-1} denominada inversa de **A**, de modo que $A^{-1}A = AA^{-1} = I$. El cálculo de esta matriz inversa se realiza mediante la siguiente fórmula:

$$A^{-1} = \frac{(Adj(A))^t}{|A|}$$

Donde **Adj(A)** denota la matriz de adjuntos de **A**. Cada celda de esta matriz contiene un valor $d_{i,j}$, que se calcula como:

$$d_{i,j} = (-1)^{i+j} \vee A_{i,j} \vee$$

Ejemplo: Cálculo de matriz inversa

$$A_{2,2} = \begin{pmatrix} 1 & 2 \\ -2 & 3 \end{pmatrix}$$

$$|A| = 1 * 3 - 2 * (-2) = 7$$

$$Adj(A) = \begin{pmatrix} 3 & 2 \\ -2 & 1 \end{pmatrix}$$

$$A^{-1} = \begin{pmatrix} 3/7 & -2/7 \\ 2/7 & 1/7 \end{pmatrix}$$

5. Estadística descriptiva

La estadística es una rama de las matemáticas que se encarga de la resolución directa de problemas de índole práctica, aplicando técnicas para la recogida de datos del problema en estudio y del uso de técnicas de análisis para la extracción de información relevante de estos datos. Dentro de un proceso de estudio estadístico podemos identificar los siguientes pasos:

1. Recogida de datos e identificación de variables estadísticas.
2. Tratamiento inicial de los datos.
3. Análisis de datos exploratorio, utilizando gráficas apropiadas.
4. Descripción de variables y características relevantes (posiblemente mediante transformación de los datos iniciales).
5. Análisis formal mediante modelos estadísticos.

La **estadística descriptiva** se puede relacionar con los puntos 2, 3 y 4 anteriores, mediante la definición de medidas estadísticas relevantes de interés y transformaciones de datos mediante agregaciones, simplificaciones, etc.

Dentro de cualquier estudio estadístico es necesario identificar las siguientes componentes:

- | **Población:** conjunto de individuos u objetos/elementos que se encuentran bajo estudio.
- | **Muestra poblacional:** Un subconjunto representativo de elementos de la población.
- | **Individuo:** Elemento particular y unívocamente identificable dentro de la población.
- | **Variable:** Característica particular, común a todos los individuos de la población, de la que se recogen datos y es susceptible de estudio.

En todo problema, trataremos con un conjunto de variables de diversa naturaleza, que serán utilizadas como variable de salida/objeto de modelado (variable dependiente), o como entrada (variable independiente/predictora). A grandes rasgos, podremos catalogar las variables según su naturaleza en los siguientes tipos:

- | **Variables cualitativas:** Son características de un individuo que aportan información no cuantificable (marca, color, forma, código postal, profesión, etc.).
- | **Variables cuantitativas discretas:** Son características de un individuo que aportan información cuantificable, pero cuyos valores son discretos (edad, fecha, número de hijos, etc.).
- | **Variables cuantitativas continuas:** Son características que aportan información cuantificable y cuya naturaleza no es discreta (altura, número de kilómetros, peso, etc.).

Cuando se realiza un proceso de recogida de muestras sobre una población, y de modo que se pueda facilitar el estudio estadístico, lo habitual es establecer resúmenes de los datos recogidos en **tablas estadísticas**, recogiendo variables y anotando sus valores y el número de veces que aparece dicho valor en la muestra (**frecuencias**). Cuando las variables de estudio tienen un carácter numérico discreto, las tablas de frecuencias se construyen para cada valor muestreado. Sin embargo, si la naturaleza de las variables es numérica continua, las tablas de frecuencias suelen resumirse por intervalos.

Ejemplo: Cálculo de tabla de frecuencias con valores discretos

Supongamos que muestreamos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3, 2, 2, 4, 3, 3, 1, 2, 3, 4}**. Construimos la tabla de frecuencias como sigue:

x_i	n_i
1	1
2	3
3	4
4	2
TOTAL:	10

En la tabla del ejemplo, el término x_i indica cada uno de los posibles valores que puede tomar la variable X en la muestra (en nuestro caso, valores {1, 2, 3, 4}). El término n_i es la frecuencia absoluta (número de individuos con valor x_i en la muestra), aunque también es común usar el término de **frecuencia relativa**, f_i , calculado como el tanto por uno de individuos en la muestra con valor x_i , o $f_i = n_i / N$.

Ejemplo: Cálculo de tabla de frecuencias con valores continuos

Supongamos que muestreemos la variable X (calificación) sobre una población de estudiantes. Tomamos un total de $N=10$ muestras, obteniendo los valores $X=\{3.5, 4.1, 6.9, 9.8, 10, 7.5, 6.1, 5.2, 2.3, 4.8\}$. Construimos la tabla de frecuencias como sigue:

x_i	n_i
[0-5)	4
[5-7)	3
[7-9)	1
[9-10]	2
TOTAL:	10

En la tabla del ejemplo, el término x_i indica cada intervalo de interés en el estudio (correspondientes a los valores categóricos *suspenso*, *aprobado*, *notable*, *sobresaliente*), que puede tomar la variable X en la muestra. El término n_i , al igual que en el caso discreto, es la frecuencia con la que aparece un valor de cada intervalo x_i en la muestra.

5.1. Medidas de posición

Las medidas de posición, o medidas estadísticas descriptivas, son un conjunto de herramientas y agregaciones de datos que permiten obtener una información resumida del comportamiento de una muestra de población. Entre ellas encontramos la media, mediana o la moda, como las más utilizadas.

La media estadística

Se calcula sobre variables numéricas, e indica el valor promedio que tienen los individuos en una característica concreta (altura promedio, número de habitaciones promedio, etc.). Siendo **X** una variable estadística numérica, y **{v_i}** los valores de **X** para un conjunto de **N** individuos que conforman una muestra de la población, la **media estadística** se calcula como:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N v_i$$

Si calculamos la media a partir de la tabla de frecuencias, la fórmula también puede reescribirse como:

$$\bar{x} = \sum_{i=1}^N f_i x_i = \frac{1}{N} \sum_{i=1}^N n_i x_i$$

Ejemplo: Cálculo de la media

Supongamos que muestreemos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3, 2, 2, 4, 3, 3, 1, 2, 3, 4}**. Sea la tabla de frecuencias obtenida la siguiente:

x_i	n_i
1	1
2	3
3	4
4	2
TOTAL:	10

Se calcula la media como:

$$\bar{X} = \frac{3 + 2 + 2 + 4 + 3 + 3 + 1 + 2 + 3 + 4}{10} = 2,7$$

O también como:

$$\bar{X} = \frac{1}{10} (1 * 1 + 3 * 2 + 4 * 3 + 2 * 4) = 2,7$$

La moda estadística

La moda estadística, aplicada sobre una variable numérica **X**, pretende representar cuál es el valor más frecuente de **X** entre todos los individuos muestreados de la población.

En el caso de **variables numéricas discretas**, analizando la tabla de frecuencias, se corresponde con aquel valor **x_i cuyo valor h_i es máximo** (número de individuos en el intervalo **x_i** dividido por la amplitud del intervalo). En el caso de **variables numéricas continuas**,

Su cálculo debe realizarse mediante la siguiente fórmula:

$$Mo(X) = e_{i-1} + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} a_i$$

Donde:

- | El valor **i** se corresponde con el índice del intervalo **x_i** cuya frecuencia es máxima.
- | El valor **a_i** se corresponde con la amplitud del intervalo (valor superior – valor inferior).
- | El valor **h_i** se calcula como **h_i=n_i/a_i**.
- | El valor **e_i** se corresponde con el extremo superior del intervalo anterior al elemento **x_i** de la moda.

Ejemplo: Cálculo de la moda en variables discretas

Supongamos que muestreamos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3, 2, 2, 4, 3, 3, 1, 2, 3, 4}**. Sea la tabla de frecuencias obtenida la siguiente:

x_i	n_i
1	1
2	3
3	4
4	2
TOTAL:	10

$$Mo(X) = \max_{x_i} \{n_i\} = 3$$

Ejemplo: Cálculo de la moda en variables continuas

Supongamos que muestreemos la variable **X** (calificación) sobre una población de estudiantes. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3.5, 6.1, 6.9, 9.8, 5.1, 7.5, 6.1, 2.2, 2.3, 4.8}**. Construimos la tabla de frecuencias como sigue:

x_i	n_i	a_i	h_i
[0-5)	4	5-0=5	4/5=0.8
[5-7)	4	7-5=2	4/2=2
[7-9)	1	9-7=2	1/2=0.5
[9-10]	1	10-9=1	1/1=1
TOTAL:	10		

En este caso el valor modal se encuentra en el intervalo [5-7) dado que tiene el máximo valor h_i , por lo que $e_i=5$, $h_i=2$, $a_i=2$.

$$Mo(X) = 5 + \frac{2 - 0.6}{(2 - 0.6) + (2 - 0.5)} 2 = 5,97$$

La mediana estadística

La mediana (**no confundir con la media**) de una variable numérica **X** se corresponde con el valor central existente para **X** en toda la muestra. En el caso de variables discretas, se calcula como:

1. Ordenar todos los elementos de **X**.
2. Si **N** (tamaño de la muestra) es impar, obtener el elemento central una vez ordenado.
3. Si **N** es par la mediana es indeterminada entre los dos valores x_i vecinos del centro. En ocasiones, se suele calcular como la media de estos dos valores.

Ejemplo: Cálculo de la mediana en variables discretas

Supongamos que muestreemos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3, 2, 2, 4, 3, 3, 1, 2, 3, 4}**. En primer lugar, ordenamos los valores:

$$X = \{1, 2, 2, 2, 3, 3, 3, 3, 4, 4\}$$

Como **N** es par, el elemento central debería caer en la posición **N/2=5**. Existe una indeterminación entre **x₅=3** y **x₆=3**. Aproximamos la mediana como la media de estos dos valores:

$$Me(X) = \frac{3 + 3}{2} = 3$$

Ejemplo: Cálculo de la mediana en variables discretas

Supongamos que muestreemos la variable **X** (edad de alumnos) sobre una población de estudiantes. Tomamos un total de **N=11** muestras, obteniendo los valores **X={30, 25, 28, 42, 37, 36, 18, 22, 31, 47, 34}**. En primer lugar, ordenamos los valores:

$$X = \{18, 22, 25, 28, 30, 31, 34, 36, 37, 42, 47\}$$

Como **N** es impar, el elemento central debería caer en la posición **(N+1)/2=6**. El elemento **x₆** vale 31:

$$Me(X) = 31$$

En el caso de variables continuas, la mediana se calcula como:

$$Me(X) = e_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} a_i$$

Donde:

- | El valor **N** es el tamaño de la muestra.
- | El valor **n_i** se corresponde con el número de elementos existentes en el intervalo **x_i**.
- | El valor **N_i** se corresponde con los valores acumulados de **n_i** desde el intervalo **x₁** hasta el intervalo **x_i** (inclusive); Es decir, $N_i = n_1 + n_2 + n_3 + \dots + n_i$.
- | El elemento central será aquel intervalo **x_i** cuyo valor **N_i** sea igual o superior a $N/2$.

Ejemplo: Cálculo de la mediana en variables continuas

Supongamos que muestreamos la variable **X** (calificación) sobre una población de estudiantes. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3.5, 6.1, 6.9, 9.8, 5.1, 7.5, 6.1, 2.2, 2.3, 4.8}**. Construimos la tabla de frecuencias como sigue:

x_i	n_i	a_i	N_i
[0-5)	4	5-0=5	4
[5-7)	4	7-5=2	8
[7-9)	1	9-7=2	9
[9-10]	1	10-9=1	10
TOTAL:	10		

En este caso el intervalo mediano es [5,7), dado que tiene su valor **N_i >= N/2** (valor $10/2=5$).

$$Me(X) = 5 + \frac{10/2 - 4}{4} 2 = 5,5$$

Percentiles

Un percentil **k**, notado como **P_k**, indica qué valor de **x_k** de una variable estadística numérica **X** deja por debajo al **k** % de los valores de la variable. Por ejemplo, **P₅₀** equivale a la mediana.

Ejemplo: Cálculo de percentiles

Supongamos que muestreamos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3, 2, 2, 4, 3, 3, 1, 2, 3, 4}**. Sea la tabla de frecuencias obtenida la siguiente:

x_i	n_i	f_i	F_i
1	1	1/10= 0.1	0.1
2	3	3/10=0.3	0.4
3	4	4/10=0.4	0.8
4	2	2/10=0.2	1.0
TOTAL:	10		

- La variable **x₁=1** recoge hasta el percentil **P₁₀**.
- La variable **x₂=2** recoge hasta el percentil **P₄₀**.
- La variable **x₃=3** recoge hasta el percentil **P₈₀**.
- La variable **x₄=4** recoge hasta el percentil **P₁₀₀**.

De especial relevancia en el análisis de datos son casos específicos de percentiles, como los denominados **cuartiles**. Un **cuartil i** se nota como **Q_i**, divide a la muestra en 4 regiones y se calcula como:

- | **$Q_1 = P_{25}$** : Valor de la variable **X** cuyo valor es superior o igual al del 25% de la muestra.
- | **$Q_2 = P_{50}$** : Valor de la variable **X** cuyo valor es superior o igual al del 50% de la muestra.
- | **$Q_3 = P_{75}$** : Valor de la variable **X** cuyo valor es superior o igual al del 75% de la muestra.
- | **$Q_4 = P_{100}$** : Valor de la variable **X** cuyo valor es superior o igual al del 100% de la muestra.

5.2. Medidas de dispersión

Las medidas de dispersión permiten resumir información sobre la homogeneidad de las observaciones realizadas sobre una muestra. Entre ellas encontramos conocidas medidas como la varianza, desviación típica, etc.

Rango

El rango de un conjunto de valores de una variable estadística **X**, notado como **R**, es la diferencia entre el máximo y mínimo valor de **X** muestreados.

Ejemplo: Cálculo del rango

Supongamos que muestreemos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **$X=\{3, 2, 2, 4, 3, 3, 1, 2, 3, 4\}$** .

El rango **$R=4-1=3$**

Varianza

La varianza trata de resumir la distancia promedio existente entre todas las muestras realizadas con respecto al valor promedio. Siendo **X** la variable estadística y **{x_i}** las mediciones realizadas sobre la misma con un total de **N** muestras, entonces se calcula como:

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

Desviación típica

La desviación típica se mide en las mismas unidades que la variable **X** sobre la que se aplica. Se calcula como la raíz cuadrada de la varianza:

$$\sigma = +\sqrt{\sigma^2}$$

Ejemplo: Cálculo de la varianza y la desviación típica

Supongamos que muestreamos la variable **X** (número de habitaciones) sobre una población de inmuebles. Tomamos un total de **N=10** muestras, obteniendo los valores **X={3, 2, 2, 4, 3, 3, 1, 2, 3, 4}**.

El cálculo de la media es:

$$\bar{x} = 2,7$$

El cálculo de la varianza es:

$$\sigma^2 = \frac{1}{9} \sum_{i=1}^{10} (x_i - \bar{x})^2 = 0,9$$

El cálculo de la desviación típica es:

$$\sigma = \sqrt{\sigma^2} = 0,95$$

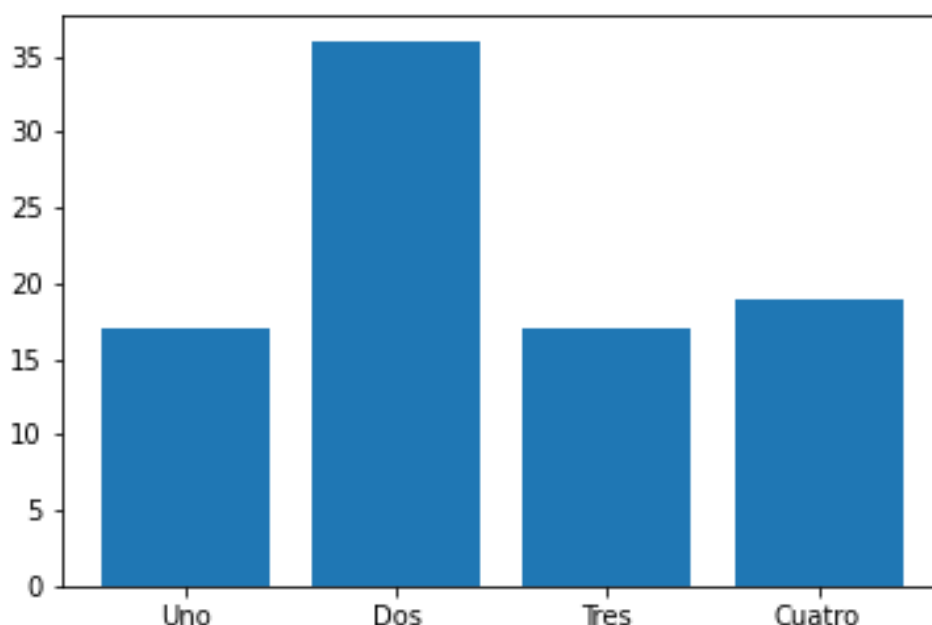
Se puede esperar que un inmueble tenga un número de habitaciones por lo normal comprendido entre $2,7-0,95= 1,75$ y $2,7+0,95=3,65$ habitaciones.

5.3. Análisis de datos exploratorio

El análisis de datos exploratorio, estudiado hoy día dentro de la rama de visualización de datos, comprende el uso de técnicas para visualizar resúmenes de los datos existentes sobre una muestra tomada de una población. Es una parte crucial dentro del análisis de datos y de la ciencia de datos y, por tanto, del Machine Learning, por lo que en este apartado revisamos algunas de las técnicas de visualización más extendidas.

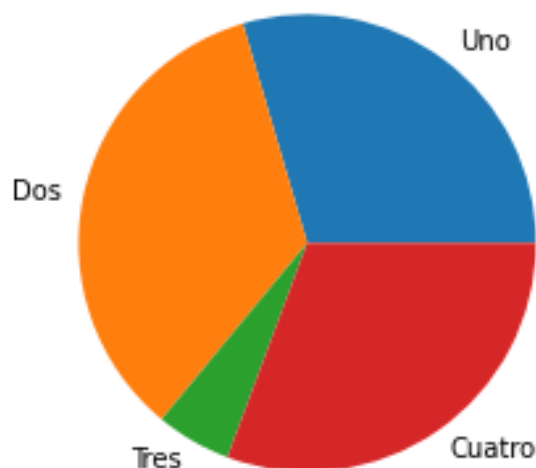
Gráficos de barras

Usados con frecuencia para mostrar resúmenes gráficos de variables cualitativas, o cuantitativas discretas. En el eje **X** se muestran cada uno de los valores de la variable, mientras que en el eje **Y** puede aparecer normalmente la frecuencia (valores n_i o f_i).



Gráficos de sectores

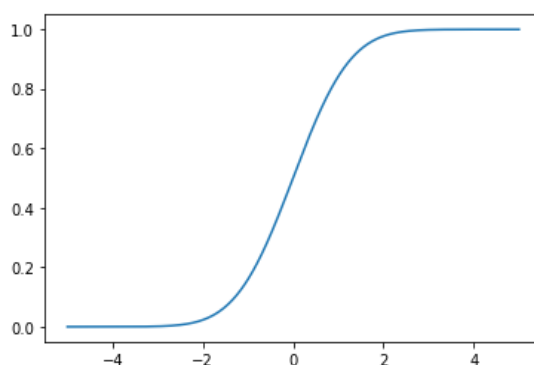
Usados también para mostrar resúmenes gráficos de variables cualitativas. Cada sector del círculo se asocia a una categoría/valor de la variable de estudio X . El área del círculo encerrada por un sector es proporcional a la frecuencia f_i .



Curvas acumulativas de distribución

Usados con frecuencia para mostrar resúmenes gráficos de variables cuantitativas discretas o continuas. En el eje X se muestran cada uno de los valores de la variable, mientras que en el eje Y puede aparecer normalmente la frecuencia acumulada (valores N_i o F_i).

En el caso continuo, el gráfico representa formalmente la integral de las frecuencias con respecto a los valores de las variables.



Histogramas

Es un caso particular de gráfico de barras. Es la forma más común de representar la distribución de la frecuencia de los datos. En el eje **X** se representan cada uno de los valores de **X** (caso discreto) o sus intervalos (caso continuo), mientras que en el eje **Y** se representa la frecuencia de cada valor.

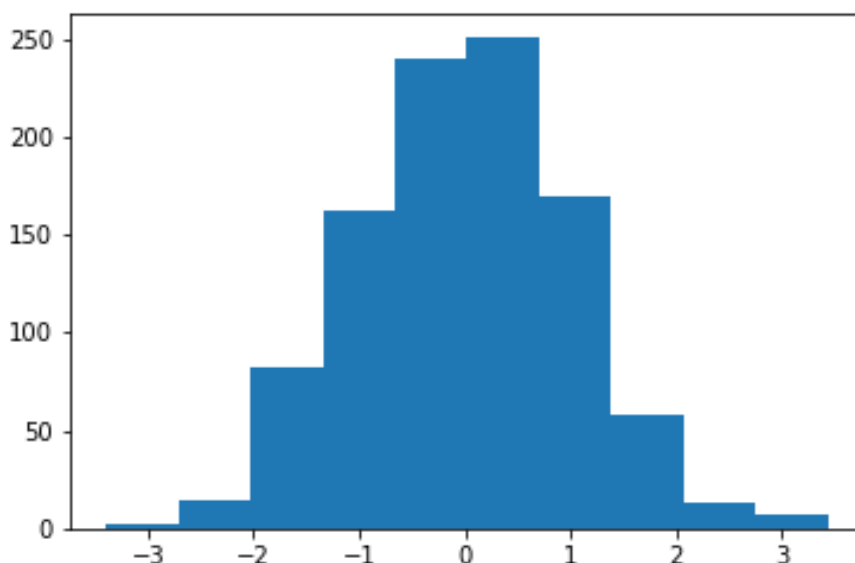
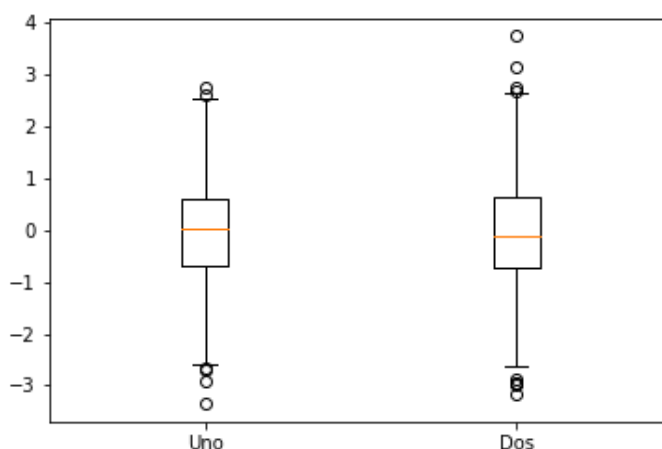


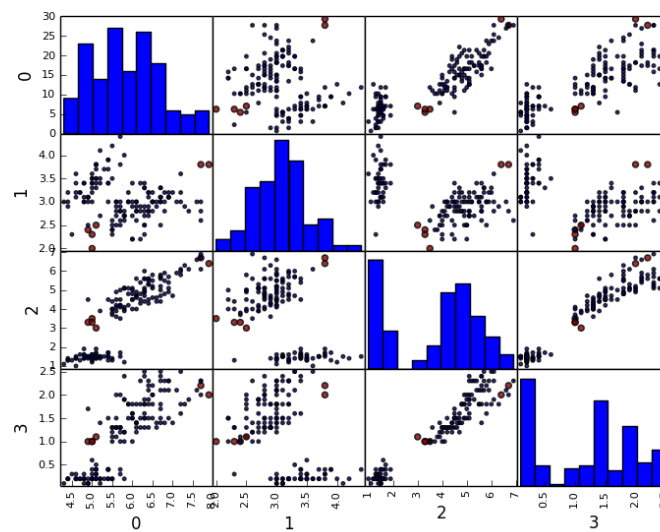
Gráfico de cajas y bigotes

Este tipo de gráficos permite representar información de la mediana de una población muestral, junto con medidas de dispersión intercuartílica (cuartiles Q_2 y Q_3), así como valores poco frecuentes u *outliers*.



Matrices dispersas (scatter matrices)

Es un tipo de gráfica compuesta muy útil, que permite reflejar gráficamente la relación existente entre pares de variables de un conjunto de datos. Suponiendo que un conjunto de datos tiene **N** variables numéricas, una **scatter matrix** está formada por **N** filas y **N** columnas. En la diagonal se muestran los histogramas de valores de cada variable, mientras que en las demás sub-figuras se muestran gráficos de dispersión por cada par de variables.



5.4. Medidas estadísticas entre varias variables

A la hora de resolver un problema, es bastante común trabajar con varias variables medidas sobre los individuos de una población (por ejemplo: edad, altura, sueldo...). En el tratamiento estadístico de los datos estaremos interesados en conocer si existe alguna relación entre estas variables o si, por el contrario, pueden considerarse independientes.

Sean dos variables estadísticas **X,Y**. Su representación en tabla de frecuencias se realiza indicando en cada fila un valor para una variable (por ejemplo **x_i**), y en cada columna el valor de la otra variable (por ejemplo **y_j**). Cada celda (**fila i, columna j**) de la tabla contiene el número de individuos que presentan simultáneamente los valores **x_i** e **y_j** , **notado como n_{ij}** .

Ejemplo: Tabla de frecuencias de 2 variables

Supongamos que muestreemos las variables discretas **X** e **Y** sobre una población de $N=8$ individuos. Obtenemos los siguientes pares (x,y) : $\{(1, 1), (1, 3), (2, 1), (2, 1), (1, 3), (2, 2), (1, 3), (1, 1)\}$

La tabla de frecuencias sería:

X\Y	1	2	3	TOTAL
1	2	0	3	5
2	2	1	0	3
TOTAL	4	1	3	8

Independencia estadística

Sean dos variables **X,Y**, muestreadas conjuntamente dando una frecuencia con valor **n_{ij}** para el número de individuos que presentan el valor **x_i** y el valor **y_j** simultáneamente. Notamos también como **$n_{i\cdot}$** a los individuos que presentan el valor **x_i** , independientemente del valor de **Y** que presenten, y a **$n_{\cdot j}$** a los individuos que toman el valor **y_j** , independientemente del valor de la variable **X** que tomen. Se dice que ambas variables son independientes, y que por tanto no se percibe relación entre ellas, si se cumple para cualquier **x_i, y_j** , que:

$$n_{ij} = \frac{n_{i\cdot} * n_{\cdot j}}{N}, \forall i, j$$

Si dos variables no son independientes, se debería estudiar la **dependencia funcional** entre ambas; es decir, si existe alguna función tal que $f(x_i) = y_j$, o que $f(y_j) = x_i$.

Covarianza

Sean dos variables **X,Y**. Se define la covarianza $Cov(X,Y)$ entre ambas variables como:

$$Cov(X,Y) = \sigma_{x,y} = \frac{\sum_i \sum_j (x_i - \bar{x})(y_j - \bar{y})}{N}$$

La covarianza nos permite conocer el comportamiento de una variable con respecto a otra. Por ejemplo, **Cov(X,Y)** es negativa cuando **X** crece e **Y** decrece; y **Cov(X,Y)** es positiva cuando ambas variables **X,Y** crecen conjuntamente. De especial interés es la matriz de varianzas y covarianzas, definida como:

$$\begin{pmatrix} \sigma_x^2 & \sigma_{x,y} \\ \sigma_{y,x} & \sigma_y^2 \end{pmatrix}$$

Coeficiente de correlación y de determinación

El coeficiente de correlación r_{xy} (o simplemente **r**) definido entre dos variables **X,Y** se calcula como:

$$r = \frac{\sigma_{x,y}}{\sigma_x \sigma_y}$$

Se utiliza para conocer cómo de correlacionadas están dos variables:

- | Si **r** está cercano a 1, entonces **X,Y** tienen una relación positiva fuerte (cuando **X** crece, entonces **Y** crece también, y viceversa).
- | Si **r** está cercano a -1, entonces **X,Y** tienen una relación negativa fuerte (cuando **X** crece, entonces **Y** decrece y viceversa).
- | Si **r** está cercano a 0, entonces no se aprecia correlación (o esta es muy débil).

Un valor **$r=1$** o **$r=-1$** puede significar dependencia funcional, de modo que una variable pueda calcularse a partir de la otra mediante una función.

De especial interés es la matriz de correlación, definida como:

$$\begin{pmatrix} r_{X,X} & r_{x,y} \\ r_{y,x} & r_{y,y} \end{pmatrix}$$

Otra medida habitualmente utilizada es el **coeficiente de determinación**, calculado como:

$$r^2 = \frac{\sigma_{x,y}^2}{\sigma_x^2 \sigma_y^2}$$

6. Conceptos básicos de probabilidad

La Teoría de la Probabilidad es una sub-rama dentro de la estadística, que formará una parte esencial dentro de algunos modelos de Inteligencia Artificial que estudiaremos en el curso.

A nivel intuitivo, nombramos la probabilidad de que ocurra un evento como un grado de certeza, medido entre 0 y 1, de que este evento tenga lugar. Al conjunto de eventos posibles lo denominaremos **variable aleatoria**.

Ejemplo: probabilidad en el lanzamiento de una moneda

Sea una variable **X** con dos posibles valores **X={cara, cruz}**, resultados posibles de lanzar una moneda al aire. La probabilidad de que salga cara la notamos como **p(X=cara)** y, si la moneda no está trucada, deberá ser **p(X=cara)=0.5**, al igual que **p(X=cruz)=0.5**.

Algunas propiedades son:

- | La probabilidad de un suceso siempre debe ser no negativa $p(X = x_i) \geq 0$.
- | La probabilidad de un suceso siempre debe ser inferior a 1 $p(X = x_i) \leq 1$.
- | La suma de las probabilidades de todos los eventos debe ser igual a 1: $\sum_i p(X = x_i) = 1$.
- | La probabilidad de que ocurra un suceso **x_i** o un suceso **x_j** se representa como $p(X = x_i \cup X = x_j) = p(X = x_i) + p(X = x_j)$.
- | La probabilidad de que ocurra un suceso **x_i** y además un suceso **x_j** se representa como $p(X = x_i \cap X = x_j)$. En el caso de que ambos sucesos sean completamente independientes, se tiene que $p(X = x_i \cap X = x_j) = p(X = x_i) * p(X = x_j)$.

Probabilidad conjunta

Cuando tenemos dos o más variables aleatorias, podemos conocer la probabilidad de que se produzcan varios eventos de forma simultánea. Por ejemplo, si tenemos dos variables aleatorias **X, Y**, se define la **probabilidad conjunta** de que se produzca el evento **x_i** de **X** y el evento **y_j** de **Y** como **$p(X=x_i, Y=y_j)$** . Por comodidad, salvo ambigüedades lo notaremos como **$p(x_i, y_j)$** .

Probabilidades marginales

Cuando tenemos dos eventos **x,y** con conocemos la probabilidad conjunta, podemos conocer la probabilidad de que se produzca de forma aislada un evento de una variable independiente de las demás. Por ejemplo, si tenemos las variables aleatorias **X,Y**, y conocemos las probabilidades conjuntas **$p(x_i, y_j)$** , podemos calcular la **probabilidad marginal $p(x_i)$** como:

$$p(x_i) = \sum_j p(x_i, y_j)$$

Igualmente, podemos calcular la **probabilidad marginal $p(y_j)$** como:

$$p(y_j) = \sum_i p(x_i, y_j)$$

Probabilidades condicionadas

Si tenemos dos sucesos **x,y**, y conocemos sus probabilidades **$p(x), p(y), p(x \cap y)$** , y **conocemos que se da uno de ellos** (por ejemplo, el evento **y**), se puede calcular el valor de probabilidad de que ocurra el evento **x** sabiendo que ya se ha producido el evento **y**. Esto se conoce como **probabilidad condicionada de x a y**, y se nota como **$p(x|y)$** . Su cálculo se realiza como:

$$p(x|y) = \frac{p(x \cap y)}{p(y)}$$

Probabilidad total

Sea un conjunto de sucesos $\{x_1, x_2, \dots, x_n\}$. A partir del conocimiento de las probabilidades condicionadas, se puede calcular la probabilidad de un suceso como:

$$p(x_j) = \sum_{i=1, i \neq j}^n p(x_i)p(x_j|x_i)$$

Teorema de Bayes

Sea un conjunto de sucesos $\{x_1, x_2, \dots, x_n\}$. El Teorema de Bayes nos permite calcular las probabilidades condicionadas de $p(x_i|x_j)$ como:

$$p(x_i|x_j) = \frac{p(x_i)p(x_j|x_i)}{p(x_j)} = \frac{p(x_i)p(x_j|x_i)}{\sum_{k=1, k \neq j}^n p(x_k)p(x_j|x_k)}$$

7.1. Distribuciones de probabilidad

Ya conocemos que un evento de una variable aleatoria discreta X puede tener un valor de probabilidad. Una **distribución de probabilidad** es una función f que asigna, a cada evento de la variable aleatoria, su valor de probabilidad. Formalmente, se define como:

$$f: X \rightarrow [0,1]$$

$$x \rightarrow f(x) = p(X = x)$$

También es interesante el concepto de **distribución de probabilidad acumulada $F(x)$** , definida en variables continuas como la integral de una función de distribución de probabilidad f .

$$F(x) = \int f(x) \partial x$$

Tanto en el caso discreto como en el continuo, la función de probabilidad acumulada equivale al cálculo de la siguiente probabilidad:

$$F(x_i) = p(x \leq x_i)$$

A continuación, estudiamos algunas distribuciones de probabilidad relevantes para el curso:

Distribución uniforme

Es el caso cuando todos los eventos presentan el mismo valor de probabilidad. Sea **X** una variable aleatoria discreta con **N** posibles eventos, entonces la función de distribución uniforme es:

$$f(x) = \frac{1}{N}$$

En el caso continuo, donde $X = [a, b]$, tenemos:

$$f(x) = \frac{1}{b - a}$$

La **notación** para la distribución uniforme continua en un intervalo **[a,b]** se escribe **U(a,b)**. De especial relevancia para nosotros sería la distribución **U(0,1)**, que indica que cualquier valor entre 0 y 1 tiene la misma probabilidad de ocurrencia.

Distribución normal

Es el caso más común y usado en estadística. Se utiliza para describir el comportamiento de múltiples problemas de la naturaleza y la ingeniería. Una distribución normal tiene dos parámetros: Media y desviación típica. La distribución normal de media **μ** y desviación típica **σ** se nota como **N(μ, σ)**, y se calcula como:

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

Al caso de la función normal $N(0,1)$ se le conoce como **distribución normal estándar**.

Cualquier distribución normal $N(\mu, \sigma)$ se puede **estandarizar**, restando a cada elemento el valor μ y dividiéndolo posteriormente por σ .

7.2. Tests de hipótesis estadísticos

En Ciencia de Datos, al igual que en otras muchas áreas (medicina, finanzas, etc.), y particularmente en el área de Machine Learning dentro de la Inteligencia Artificial, es necesario realizar pruebas que corroboren las hipótesis de un experimento con respecto a los resultados obtenidos. Usualmente, esto se realiza mediante la ejecución de tests de hipótesis sobre los datos y los resultados que se tienen. En esta sección revisamos algunos tipos de tests de hipótesis ampliamente utilizados a lo largo del curso. Todos los tests estadísticos darán como resultado (entre otras cosas) un valor de probabilidad (o *p-value*) que indique si se acepta la hipótesis planteada en el test usado en particular, o si se rechaza. El hecho de aceptar o rechazar una hipótesis vendrá dada por un **nivel de confianza** que se establece a priori. Por ejemplo, en el área de la medicina se establece un nivel de confianza estándar del 99%, y en Ciencias de la Computación se suele usar un nivel de confianza estándar del 95%.

Tests de normalidad

Se utilizan para corroborar que una muestra poblacional puede ser modelada mediante una distribución normal. Ejemplos de este tipo de tests son **Shapiro-Wilk** o **D'Agostino's K-squared test**.

Si un conjunto de datos pasa el test de normalidad, entonces **se puede asumir que la media y la desviación típica son medidas estadísticas que reflejan fielmente el comportamiento de la población.**

En Ciencia de Datos, los tests de normalidad suelen ser los primeros que se aplican sobre los resultados. Si un conjunto de datos pasa un test de normalidad, podremos hacer uso de **tests paramétricos** para comparación y validación de resultados. En caso contrario, la media y desviación típica no reflejan fielmente el comportamiento de la población y deberemos establecer otros estadísticos de resumen (mediana, moda, ...), y aplicar **tests no paramétricos de comparación**.

Tests paramétricos: t-Student de dos muestras

El *T-test* de dos muestras o test de Student de dos muestras, suele usarse para comprobar si dos distribuciones de datos (asumidas normales) tienen la misma media y, por tanto, son similares. Usaremos este tipo de tests para validar resultados y comparar métodos de aprendizaje en módulos posteriores, cuando los datos que obtengamos sean normales.

Tests no paramétricos: Wilcoxon, Kruskal-Wallis

Cuando las distribuciones de datos no pasan un test de normalidad, se debe usar un test no paramétrico para comparar dos poblaciones de resultados. En este caso, usaremos el test de Wilcoxon para comparar dos poblaciones, o el test de Kruskal-Wallis para comparar múltiples poblaciones.

7. Métodos de Monte-Carlo

La simulación por Monte-Carlo es un método muy simple, a la vez que potente, para modelar sistemas o problemas que tienen componentes aleatorias/probabilísticas. Por ejemplo: El modelado de un sistema de organización ferroviario, el control y previsión de inventarios, sistemas de atención al cliente en supermercados, etc.

El método de Montecarlo se utiliza con bastante frecuencia en ciertas áreas dentro de la Inteligencia Artificial, entre las que se citan la simulación y el modelado de sistemas, o el aprendizaje por refuerzo. Está compuesto por los siguientes pasos:

1. Identificar las variables aleatorias y su relación con el resto de variables dentro del problema.
2. Elaborar distribuciones de probabilidad acumuladas para cada una de las variables seleccionadas en el punto anterior.
3. Generación de números aleatorios en el rango de cada variable.
4. Simular el sistema y obtener resultados.

Ejemplo: Simulación de inventario de vacunas diarias

En un centro de salud se dispone de un servicio de vacunación para un tipo de vacuna con fecha de consumo inmediata. De este modo, se debe adquirir diariamente un número determinado de vacunas a administrar ese día. A fin de poder optimizar los costes de adquisición de las vacunas con el proveedor, se desea realizar una estimación del número de vacunas diario que deberían adquirirse.

Se ha realizado un experimento a lo largo de 30 días, observando cuántas vacunas ha sido solicitadas.

Vacunas

N.º de días

0	4
1	3
2	3
3	4
4	2
5	2
6	3
7	2
8	5
9	2

Cálculo de probabilidades y distribución de probabilidad acumulada:

Vacunas

Prob.

**Prob.
acumulada**

0	4/30	4/30
1	3/30	7/30
2	3/30	10/30
3	4/30	14/30
4	2/30	16/30
5	2/30	18/30
6	3/30	21/30
7	2/30	23/30
8	5/30	28/30
9	2/30	1.0

A continuación, realizaremos un experimento muestreando 10 valores al azar, generando números aleatorios entre 1 y 30. Se selecciona el número de vacunas asignado a cada fila de la probabilidad acumulada asociada. Supongamos que los valores aleatorios son los números: 5, 23, 10, 18, 2, 14, 14, 19, 22, 20. Entonces, el número de vacunas seleccionado es, en cada instancia del experimento:

Valor aleatorio	Vacunas
5	1
23	7
10	2
18	5
2	0
14	3
14	3
19	6
22	7
20	6

Por tanto, la demanda esperada de vacunas mediante la simulación sería la suma de los valores de vacunas resultados de la simulación, dividido por el total de experimentos:

$$\frac{1+7+2+5+0+3+3+6+7+6}{10} = 4 \text{ vacunas.}$$

Si calculamos el valor esperado mediante la media, obtenemos el valor:

$$\overline{\text{vacuna}} = \frac{4 \cdot 0 + 3 \cdot 1 + 3 \cdot 2 + 4 \cdot 3 + 2 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 + 2 \cdot 7 + 5 \cdot 8 + 2 \cdot 9}{30} = 4,3$$

En ejemplos tan sencillos, el valor simulado cuando crece el número de experimentos debe aproximarse al valor esperado (media).

8. Nociones de estadística para clasificación binaria

Uno de los grandes problemas con los que trata la Inteligencia Artificial es el problema de clasificación. Este problema consiste, esencialmente, en encontrar a qué categoría pertenece una muestra recién adquirida. Ejemplos de aplicación son el reconocimiento de objetos en imágenes, la clasificación de calidad de verduras con respecto a sus características, la previsión de concesión de préstamos hipotecarios, categorización de clientes, etc.

La **clasificación binaria** o **dicotómica** es un caso particular del problema de clasificación, donde sólo existe una (o dos) clases a diferenciar, como por ejemplo: detección de intrusos en viviendas, detección de correo SPAM o, en general, una pregunta cuya respuesta pueda ser expresada en términos de "Sí" o "No" (usualmente, usaremos los términos "*positivo*" y "*negativo*").

Un **clasificador** es un modelo matemático que permite, a partir de datos de variables denominadas **atributos** o **variables independientes**, predecir un valor o **clase/variable dependiente** asociada a dichos atributos. En argot del aprendizaje automático, también se conocen como **patrones de entrada (variables independientes) y salida (variable dependiente)**. También se suele usar el término **patrón** para referirnos al par (atributos, clase) de un individuo de la población.

Algunos ejemplos de problemas de clasificación son: Determinar si un estudiante es bueno o no en función de sus calificaciones, identificar si el agua contiene algún residuo nocivo para la salud partiendo de sus análisis clínicos, conocer si un paciente sufre cáncer o no a partir de sus análisis clínicos, reconocer un dígito a partir de una imagen de texto manuscrito, etc.

De momento no entraremos en detalle sobre cómo se construye un clasificador, aunque nos centraremos en clasificadores binarios. Modelaremos un clasificador binario como una función $f: X \rightarrow \{\text{positivo}, \text{negativo}\}$ que, dado como entrada un conjunto de valores de atributos \mathbf{x} , proporciona el valor $f(\mathbf{x})=\text{positivo}$ si los atributos se reconocen y se asocian a una clase, y el valor $f(\mathbf{x})=\text{negativo}$ si se asocian a otra clase.

La principal tarea abordada cuando se soluciona un problema de clasificación reside en ajustar el clasificador para que funcione de forma óptima. Una vez que se ajusta dicho clasificador, debemos comprobar su correcto funcionamiento. Desde el punto de vista estadístico existen múltiples medidas para realizar dicha tarea, aunque comentamos las más comunes:

| **Positivos bien clasificados (*True Positive, TP*)**. Indica el número de patrones $\{\mathbf{x}_i\}$ en la muestra que deben tener clase positiva y el clasificador devuelve $f(\mathbf{x}_i)=\text{positivo}$. Cuando el valor se expresa en tanto por uno, se denomina **tasa de positivos bien clasificados (*True Positive Rate, TPR*)**.

$$TPR = \frac{TP}{TP + FN}$$

| **Falsos positivos (*False Positive, FP*)**. Indica el número de patrones $\{\mathbf{x}_i\}$ en la muestra que deben tener clase negativa y el clasificador devuelve $f(\mathbf{x}_i)=\text{positivo}$. Cuando el valor se expresa en tanto por uno, se denomina **tasa de falso positivos (*False Positive Rate, FPR*)**.

$$FPR = 1 - TPR$$

| **Negativos bien clasificados (*True Negative, TN*)**. Indica el número de patrones $\{\mathbf{x}_i\}$ en la muestra que deben tener clase negativa y el clasificador devuelve $f(\mathbf{x}_i)=\text{negativo}$. Cuando el valor se expresa en tanto por uno, se denomina **tasa de negativos bien clasificados (*True Negative Rate, TNR*)**.

$$TNR = \frac{TN}{TN + FP}$$

Falsos negativos (*False Negative, FN*). Indica el número de patrones $\{x_i\}$ en la muestra que deben tener clase positiva y el clasificador devuelve **$f(x_i)$ =negativo**. Cuando el valor se expresa en tanto por uno, se denomina **tasa de negativos bien clasificados (*False Negative Rate, FNR*)**.

$$FNR = 1 - TNR$$

Precisión (*accuracy*): Es la proporción:

$$acc = \frac{TP + TN}{N}$$

Donde **N** es el tamaño de la muestra.

Tablas de confusión

Una tabla de confusión permite visualizar tabularmente un resumen del comportamiento de un clasificador. Ayuda, fundamentalmente, a comprobar la precisión del clasificador y si éste tiene algún tipo de sesgo (por ejemplo, alto porcentaje de FP/FN).

La tabla de confusión se construye como sigue: En las filas y columnas se colocan los distintos valores de la clase. Cada fila se asocia a un valor de un patrón, y cada columna se asocia a un valor devuelto por el clasificador. En el problema de clasificación binaria, la tabla de confusión tiene sólo dos filas y dos columnas, y sus celdas contienen la siguiente información:

		Clasificador	
		Positivo	Negativo
Datos reales	Positivo	TP	FN
	Negativo	FP	TN

Por ejemplo, la fila "negativo" y columna "positivo" debe contener el valor **FP**, indicando el número de patrones en la muestra que deben ser clasificados como negativos, pero que el clasificador indica que son positivos. Un clasificador perfecto tendría sólo elementos en la diagonal (en clasificación binario, esto significaría que no hay FP ni FN).

Curvas ROC y AUC

Las curvas ROC/AUC se utilizan para evaluar visualmente el comportamiento de un clasificador binario. La curva ROC (Receiver Operating Characteristic) muestra, en el eje X, la evolución de la tasa de falsos positivos, y en el eje Y la tasa de positivos bien clasificados. Se genera calculando diferentes umbrales para los que el clasificador debe devolver positivo o negativo, y mostrando los valores FPR y TPR en los ejes de coordenadas.

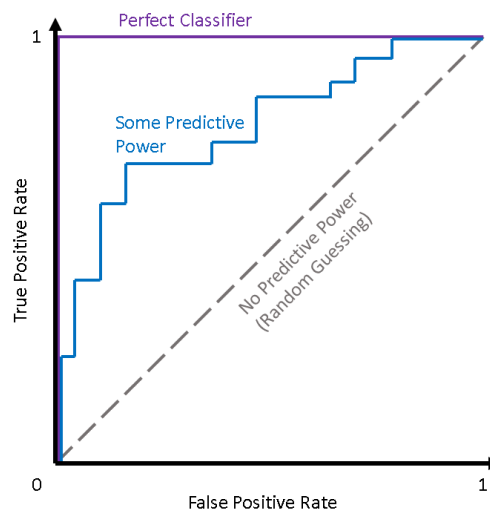


Imagen obtenida de: <https://polmartisanahuja.com/entendiendo-la-curva-roc-y-el-auc-dos-medidas-del-rendimiento-de-un-clasificador-binario-que-van-de-la-mano/>

La curva ROC se utiliza fundamentalmente para calcular el valor **AUC (Area Under Curve)**, que se corresponde con la integral de la gráfica **ROC**. El valor **AUC** permite conocer cómo de bueno es el comportamiento de un clasificador, siendo el valor **AUC** de un clasificador óptimo igual a 1.

vuela

**PLATAFORMA
DIGITAL DE
ANDALUCÍA**

