

MATEMÁTICAS PARA LA INTELIGENCIA ARTIFICIAL



**MÁSTER UNIVERSITARIO EN INTELIGENCIA
ARTIFICIAL**



Universidad
Internacional
de Valencia

Este material es de uso exclusivo para los alumnos de la Universidad Internacional de Valencia. No está permitida la reproducción total o parcial de su contenido ni su tratamiento por cualquier método por aquellas personas que no acrediten su relación con la Universidad Internacional de Valencia, sin autorización expresa de la misma.

Edita

Universidad Internacional de Valencia

Máster Universitario en
Inteligencia Artificial

Matemáticas para la inteligencia artificial
6 ECTS

Leyendas



Enlace de interés



Ejemplo



Importante



abc Los términos resaltados a lo largo del contenido en color **naranja** se recogen en el apartado **GLOSARIO**.

Contenido

CAPÍTULO 1. LÓGICA	7
1.1. Lógica proposicional.....	7
1.1.1. Enunciados lógicos y valores de verdad	8
1.1.2. Conectores lógicos	9
1.2. Álgebra de Boole	17
1.2.1. Definición y ejemplos	17
1.2.2. Propiedades y resultados	19
1.3. Funciones booleanas y puertas lógicas	20
1.3.1. Funciones booleanas.....	20
1.3.2. Puertas lógicas	22
1.3.3. Circuitos lógicos	27
1.4. Lógica de predicados	30
1.4.1. Predicados y universos	31
1.4.2. Cuantificadores	32
1.4.3. Semántica y propiedades	35
1.4.4. Cálculo de predicados.....	35
1.5. Lógica inductiva.....	41
1.5.1. Programación lógica inductiva: definición y ejemplos	41
1.5.2. Búsqueda de hipótesis.....	44
1.5.3. Inducción predictiva y descriptiva.....	47
1.6. Teoría de complejidad computacional	48
1.6.1. Definición de complejidad computacional: uso en el diseño de algoritmos.....	49
1.6.2. Modelos de computación.....	50
1.6.3. Clases de complejidad.....	51
1.7. Teoría de grafos.....	51
CAPÍTULO 2. ÁLGEBRA	74
2.1. Teoría de conjuntos.....	74
2.1.1. Operaciones entre conjuntos	75
2.1.2. Partes de un conjunto	78
2.2. Relaciones y aplicaciones.....	79
2.2.1. Relaciones entre conjuntos	80
2.2.2. Relaciones de equivalencia.....	81

2.2.3. Aplicaciones entre conjuntos	83
2.3. Combinatoria	90
2.3.1. Principios de cardinalidad	90
2.3.2. Permutaciones, combinaciones y variaciones	93
2.4. Espacios vectoriales	96
2.4.1. Subespacios vectoriales	100
2.4.2. Sistemas libres, generadores y bases	102
2.4.3. Aplicaciones lineales	104
2.5. Factorización de matrices	118
2.5.1. Diagonalización	119
2.5.2. Factorización <i>LU</i>	126
2.5.3. Factorización de Cholesky	126
2.5.4. Factorización QR	126
2.5.5. Descomposición en valores singulares (SVD)	127
2.6. Cálculo tensorial	127
2.6.1. Concepto de rango y dimensión de un tensor	128
2.6.2. Tensores covariantes y contravariantes	129
2.6.3. Introducción a TensorFlow	130
CAPÍTULO 3. ANÁLISIS	133
3.1. Cálculo infinitesimal	134
3.1.1. Límites	134
3.1.2. Derivadas de funciones	136
3.1.3. Crecimiento, extremos relativos y extremos absolutos	139
3.1.4. Máximos, mínimos y puntos críticos: optimización	141
3.1.5. Primitivas e integral definida	144
3.2. Cálculo multivariable	147
3.2.1. Derivadas parciales	148
3.2.2. Gradiente y derivadas direccionales	149
3.2.3. Segundas derivadas parciales y matriz hessiana	150
3.2.4. Máximos, mínimos y puntos críticos: optimización	150
3.3. Algoritmos de gradient descent y forward & backpropagation	151
3.3.1. Gradient descent	151
3.3.2. Forward & backpropagation	152
3.3.3. Aplicación y ejemplos con TensorFlow	154
Resumen	156

CAPÍTULO 4. PROBABILIDAD Y ESTADÍSTICA	157
4.1. Introducción a la probabilidad	158
4.1.1. Independencia de sucesos	160
4.1.2. Probabilidad condicionada y teorema de Bayes	162
4.1.3. Variables aleatorias discretas y continuas	164
4.1.4. Esperanza y varianza de una variable aleatoria	165
4.1.5. Distribuciones discretas y continuas	169
4.2. Estimación de parámetros	174
4.2.1. Límites de variables aleatorias.....	175
4.2.2. Método de los momentos	176
4.2.3. Método de la máxima verosimilitud	177
4.3. Procesos estocásticos.....	179
4.3.1. Matrices estocásticas.....	179
4.3.2. Cadenas de Markov.....	180
4.3.3. Procesos gaussianos	181
4.4. Desarrollo de algoritmos probabilísticos con TensorFlow	182
GLOSARIO	187
ENLACES DE INTERÉS	197
BIBLIOGRAFÍA.....	198



Capítulo 1

Lógica

Introducción

La Inteligencia Artificial es una cualidad atribuida a las máquinas por la que, a través de diferentes mecanismos, estas demuestran tener una inteligencia automática. Este concepto hace aparecer el de “machine learning” o aprendizaje automático, el cual está, a su vez, relacionado con los algoritmos que se utilizan para la programación de la máquina, de tal manera que esta sea capaz de predecir su comportamiento futuro adecuado en base a unos conceptos que ya se le han incluido y que le permiten, además, recabar información de las situaciones reales a las que se enfrente.

Para poder realizar esta programación es necesario partir de un conocimiento previo de las matemáticas, conocimiento que, en primer lugar, está relacionado con la lógica.

En matemáticas, la lógica es entendida como el conjunto de teorías y paradigmas que llevan a distinguir entre lo que es correcto (o verdad) y lo que es incorrecto (o falso). Estos procesos se centran en un problema o hipótesis que debe resolverse utilizando diferentes mecanismos. Cada una de las formas de acercarse a esta solución dará lugar a un tipo de proceso lógico.

Es importante tener en cuenta, durante la lectura de las siguientes páginas y en las posibles consultas, que el trabajo relacionado con la Inteligencia Artificial es relativamente moderno, por lo que los autores españoles manejan material de base inglesa, lo que puede llevar a encontrar el mismo concepto definido con varios términos diferentes, siendo todos ellos correctos. No obstante, y en este sentido, será el mismo contexto el que puede aclarar el significado del concepto trabajado.

Objetivos

- Conocer las aplicaciones de la lógica matemática para su desarrollo en la Inteligencia Artificial.
- Manejar los diferentes tipos de lógica, desde la proposicional a la teoría de complejidad computacional.
- Plantear los algoritmos necesarios según el objetivo marcado para la máquina.

1.1. Lógica proposicional

La lógica proposicional está presente de forma implícita o explícita en una gran variedad de construcciones de carácter lingüístico, en las que se establecen conexiones entre diferentes enunciados o relaciones de causalidad entre estos. Dichas conexiones y relaciones pueden formalizarse hasta constituir los fundamentos de esta teoría, que será el objeto de estudio en el capítulo presente.

El estudio de los enunciados y sus valores de verdad a través de la lógica proposicional constituye un caso particular de álgebra de Boole, por lo que ahora nos ocuparemos en profundizar más en lo relativo a ésta y sus propiedades. Asimismo, utilizaremos propiedades de dicha álgebra para establecer estrategias para simplificar afirmaciones lógicas complejas.

En lo que respecta al contenido de este tema, el lector interesado puede complementar la información en libros como los de Castel de Haro y Llorens (*Lógica de primer orden*), Manzano y Sánchez (*Lógica para principiantes*) y Miranda (*Lógica y métodos discretos*).

1.1.1. Enunciados lógicos y valores de verdad

Un **enunciado lógico** o **proposición lógica** consiste en cualquier afirmación a la cual se le pueda atribuir un **valor de verdad**, pudiendo ser este verdadero (V) o falso (F). Obsérvese que puede realizarse una identificación con los elementos del conjunto $B = \{0, 1\}$, siendo 1 el valor asociado a un enunciado verdadero, V, y 0 el valor asociado a un enunciado falso, F.

Hay una amplísima variedad de afirmaciones que realizamos cada día en conversaciones o debates cotidianos que están sujetas a la asignación de valores de verdad. Se muestran a continuación algunos ejemplos:

- Juan se ha gastado 40 € en un videojuego.
- María llevó ayer el coche al taller.
- Las patatas que compraste el otro día estaban rancias.
- Aquella serie que me recomendaste el mes pasado es muy buena.

Podemos observar a partir de las frases anteriores que existe una sutil diferencia entre el grupo formado por las dos primeras y las dos últimas. En el primer caso, puede atribuirse un valor de verdad, sin dar lugar a ambigüedades, si los hechos que describen las afirmaciones se realizaron según lo descrito: en el primer caso, si Juan realmente se gastó 40 € en un videojuego y si María realmente llevó el coche al taller el día antes a la realización de la afirmación. Por el contrario, en el segundo caso las afirmaciones implican juicios de valor subjetivos, a los cuales es discutible asignar un valor de verdad, en tanto que emplean conceptos ambiguos, como “estar rancias” o “ser muy buena”.

Nos centraremos en el primer tipo de afirmaciones, es decir, aquellas en las que se pueda establecer un valor de verdad de forma objetiva, o sea, sin ambigüedades. Tal y como se ha indicado anteriormente, esto implica que solo pueden tener uno de los dos posibles valores de verdad (V/F), sujeto a la previa comprobación de su veracidad o falsedad, respectivamente.

Dependiendo del tipo de proposición lógica considerada, pueden clasificarse en tres tipos: tautología, contradicción o contingencia, los cuales se detallan a continuación.

a) Tautología

Una proposición **p** lógica es tautológica si siempre es verdadera, independientemente de la situación o el contexto considerado.

Presentamos a continuación una serie de enunciados tautológicos, con grado de complejidad variable.

- Todo múltiplo de 6 es también múltiplo de 3.
- Mañana lloverá, o quizás no lo haga.
- Existen infinitos números naturales.
- Si lanzamos un dado, puede que salga un número par o bien uno impar.
- Mi equipo de fútbol encabeza la liga a 4 puntos del segundo y a falta de disputarse una jornada, por lo que ya la han ganado.

b) Contradicción

Una proposición lógica **p** es contradictoria si siempre es falsa, independientemente de la situación o contexto considerado.

A continuación, se muestran una serie de enunciados contradictorios de diferente complejidad.

- 10 es un número primo.
- El rascacielos mide un total de metros, es decir, kilómetros.
- No existe un número infinito de números naturales.
- Ayer escuchó un total de 3 canciones de un grupo y 2 de otro, por lo que escuchó un total de 8 canciones.
- El siguiente enunciado es verdadero: el enunciado anterior es falso.

c) Contingencia

Una proposición lógica **p** es contingente si puede ser verdadera o bien falsa dependiendo de la situación o el contexto considerado.

Existe una amplia variedad de enunciados contingentes. A continuación, se muestran algunos ejemplos:

- Mañana lloverá.

- Si lanzamos una moneda, saldrá cara.
- Hoy es martes.
- Si elegimos un dígito al azar, será mayor que 3.
- Si un conjunto tiene elementos y otro solo , entonces el último está incluido en el primero.

1.1.2. Conectores lógicos

En los ejemplos anteriores, hemos visto que se han empleado enunciados lógicos de diferente complejidad, involucrando en ocasiones combinaciones de afirmaciones que forman un todo. Son destacables especialmente los conectores empleados para unirlas, fundamentalmente la negación “no”, la conjunción “y”, la disyunción “o” y el condicional “entonces”.

1.1.2.1. Definiciones y ejemplos

En esta sección, estudiaremos a fondo dichos conectores, algunos ejemplos asociados y su relación con algunas de las álgebras de Boole.

Disyunción

Dados dos enunciados lógicos p y q , se define como disyunción entre p y q un nuevo enunciado lógico, denotado por $p \vee q$ (y leído como “ p o q ”), cuyo valor de verdad está en función de los dos enunciados iniciales, cumpliéndose que la proposición lógica resultante es verdadera siempre y cuando alguno de los dos enunciados originales (o ambos) son verdaderos.

De la definición se obtiene que la tabla de verdad correspondiente a la disyunción es la siguiente:

p	q	$p \vee q$
F	F	F
F	V	V
V	F	V
V	V	V

Obsérvese que el comportamiento de este conector es el mismo que el de la puerta lógica OR, con la correspondiente asignación de los valores de verdad verdadero (V) y falso (F) a los valores binarios en B0 y 1, respectivamente.



Ejemplo

- Tomemos las afirmaciones p “5 es múltiplo de ”2 (falsa) y q “5 es múltiplo de 3” (falsa). Entonces, la afirmación $p \vee q$ “5 es múltiplo de 2 o bien múltiplo de 3” es también falsa por serlo ambas.
- Sean las afirmaciones p “4 es múltiplo de ”2 (verdadera) y q “4 es múltiplo de 3” (falsa). Entonces, la afirmación $p \vee q$ “4 es múltiplo de 2 o bien múltiplo de 3” es verdadera por haber una de ellas (p) que lo es.
- Consideremos las afirmaciones p “9 es múltiplo de 2” (falsa) y q “9 es múltiplo de 3” (verdadera). Entonces, la afirmación $p \vee q$ “9 es múltiplo de 2 o bien múltiplo de 3” es verdadera por haber una de ellas (q) que lo es.
- Tomamos las afirmaciones p “6 es múltiplo de 2” (verdadera) y q “6 es múltiplo de 3” (verdadera). Entonces la afirmación $p \vee q$ “6 es múltiplo de 2 o bien múltiplo de 3” es verdadera por serlo ambas.

Conjunción

Dados dos enunciados lógicos p y q , se denomina conjunción entre p y q a un nuevo enunciado lógico, denotado por $p \wedge q$ (y leído como “ p y q ”), cuyo valor de verdad está en función de los dos enunciados originales, de forma que el resultado es únicamente verdadero si tanto p como q lo son simultáneamente.

De acuerdo con la definición, la tabla de verdad asociada a la conjunción se corresponde con la que se muestra a continuación:

p	q	$p \wedge q$
F	F	F
F	V	F
V	F	F
V	V	V

Nótese que, en esta ocasión, el comportamiento de este conector lógico puede identificarse con el de la puerta lógica AND, asociando los correspondientes valores de verdad a sus respectivos valores binarios asociados.



Ejemplo

- Tomemos las afirmaciones p “5 es múltiplo de 2” (falsa) y q “5 es múltiplo de 3” (falsa). Entonces, la afirmación $p \wedge q$ “5 es simultáneamente múltiplo de 2 y múltiplo de 3” es también falsa por serlo ambas.
- Sean las afirmaciones p “4 es múltiplo de 2” (verdadera) y q “4 es múltiplo de 3” (falsa). Entonces, la afirmación $p \wedge q$ “4 es simultáneamente múltiplo de 2 y múltiplo de 3” es falsa por haber una de ellas (q) que lo es.
- Consideremos las afirmaciones p “9 es múltiplo de 2” (falsa) y q “9 es múltiplo de 3” (verdadera). Entonces, la afirmación $p \wedge q$ “9 es simultáneamente múltiplo de 2 y múltiplo de 3” es falsa por haber una de ellas (p) que lo es.
- Tomamos la afirmación p “6 es múltiplo de 2” (verdadera) y q “6 es múltiplo de 3” (verdadera). Entonces la afirmación $p \wedge q$ “6 es simultáneamente múltiplo de 2 y múltiplo de 3” es verdadera por serlo ambas.

Puede comprobarse que el hecho de que se cumplan simultáneamente p (“ser múltiplo de 2”) y q (“ser múltiplo de 3”) es equivalente a la afirmación “ser múltiplo de 6”. Por tanto, el enunciado lógico $p \wedge q$ es “ser múltiplo de 6”.

Condicional

Sean dos enunciados lógicos p y q . Se define el condicional o **implicación lógica** p implica q , denotado por $p \rightarrow q$ (y leído como “ p implica q ”), a un enunciado lógico cuyo valor de verdad depende de los dos enunciados originales, de forma que si se da p , entonces necesariamente debe darse también q . El primer miembro, p , recibe el nombre de causa y el segundo miembro, q , se denomina consecuencia.

La tabla de la verdad asociada, teniendo en cuenta la descripción anterior, presenta el siguiente aspecto:

p	q	$p \rightarrow q$
F	F	F
F	V	V
V	F	V
V	V	V

Se observa que para este conector lógico, el único caso en el que es falso es aquel en el que se contradice la propia definición (si ocurre p , entonces necesariamente debe ocurrir q), es decir, cuando se tiene que ocurre p pero no ocurre q .



Ejemplo

- Consideremos la afirmación ***p*** “los burros vuelan” y la afirmación ***q*** “ $2+2 = 5$ ”. En este caso, está claro que ambas afirmaciones son falsas, por lo que $p \rightarrow q$, es decir, “si los burros vuelan, entonces $2+2 = 5$ ” es verdadera. En este caso, como ***p*** es falsa, la relación de causalidad $p \rightarrow q$ no obliga al cumplimiento de ***q*** (puesto que no se da ***p***), que también es falsa, por lo que dicha implicación lógica es válida.
- Tomemos ahora la afirmación ***p*** “los burros vuelan” y la afirmación ***q*** “ $2+2 = 4$ ”. En este caso, ***p*** es falsa mientras que ***q*** es verdadera, siendo el enunciado lógico $p \rightarrow q$ equivalente a la afirmación “si los burros vuelan, entonces $2+2 = 4$ ”. Nuevamente el resultado arroja un enunciado lógico verdadero, puesto que no se cumple ***p*** y, por tanto, no necesariamente se ha de dar la relación de causalidad establecida (si bien en este caso ***q*** es verdadero, por lo que, si ***p*** hubiese sido verdadero, $p \rightarrow q$ también habría sido verdadero, igualmente).
- Consideremos los enunciados ***p*** “la Tierra no es plana” y ***q*** “ $3 < 2$ ”. En este caso ***p*** es verdadero, por lo que el enunciado $p \rightarrow q$, “si la Tierra no es plana, entonces $3 > 2$ ” es falso por serlo ***q***, ya que la veracidad de ***p*** unida a la relación de causalidad “***p*** implica ***q***” obliga a que ***q*** tenga que ser verdadero también, lo cual no es cierto.
- Por último, tomamos ***p*** “la Tierra no es plana” y ***q*** “ $3 > 2$ ”. Ahora se tiene que tanto ***p*** como ***q*** son enunciados lógicos verdaderos, por lo que la relación $p \rightarrow q$, es decir, “si la Tierra no es plana, entonces $3 > 2$ ”, es cierta, ya que el cumplimiento de ***q*** ha de implicar el cumplimiento de ***p***, que también se da, otorgándole por tanto la veracidad a la implicación lógica.

No debe confundirse la implicación lógica con la equivalencia (que veremos a continuación). Es decir, el enunciado $p \rightarrow q$ (si ***p*** se cumple, entonces ***q*** también debe verificarse) **no implica**, en general, que se cumpla también el enunciado **recíproco $p \rightarrow q$** (si ***q*** se cumple, entonces ***p*** también debe verificarse). Esto se debe a que las tablas de la verdad de $p \rightarrow q$ permiten que ***q*** pueda cumplirse sin necesidad de que ***p*** lo haga.

Por ejemplo, tomemos ***p*** “ser múltiplo de 4” y ***q*** “ser múltiplo de 2”. Naturalmente, todo múltiplo de 4 es también un múltiplo de 2, por lo que se da la relación $p \rightarrow q$. No obstante, la relación recíproca, $q \rightarrow p$ es falso, ya que existen múltiplos de 2 (***q*** verdadero) que no son múltiplos de 4 (***p*** falso): por ejemplo, el número 6, que es múltiplo de 2 pero no de 4.

Equivalencia

Sean ***p*** y ***q*** y dos enunciados lógicos. Se define como equivalencia, **bicondicional** o **implicación doble** a un nuevo enunciado lógico, denotado como $p \leftrightarrow q$ (y leído como “***p*** si y solo si ***q***”), cuyo valor de verdad depende nuevamente de ***p*** y de ***q***, de forma que únicamente es verdadero cuando el valor de verdad de ambos enunciados iniciales es el mismo.

Por tanto, la tabla de verdad asociada a este conector presenta el aspecto que se muestra a continuación:

p	q	$p \leftrightarrow q$
F	F	F
F	V	V
V	F	V
V	V	V

Es decir, en este caso, la conexión entre ambos enunciados solo es verdadera cuando comparten el mismo valor, de ahí el nombre de equivalencia, ya que dicha relación acarrea que sean equivalentes como enunciados lógicos.

Ejemplo

Si tomamos como **p** el enunciado “ser simultáneamente múltiplo de 2 y de 3” y como **q** el enunciado “ser múltiplo de 6”, entonces $p \leftrightarrow q$ se cumple por ser ambos enunciados equivalentes. Por el contrario, los enunciados **p** “ser simultáneamente múltiplo de 2 y de 3” y “**q** ser múltiplo de 9” no son equivalentes, por lo que no se cumple $p \leftrightarrow q$, puesto que, por ejemplo, 6 es un número que verifica **p** pero no verifica **q**.

Negación lógica

Sea **p** un enunciado lógico. Se define como negación lógica a un nuevo enunciado resultante de invertir el valor de la verdad del enunciado original, denotándose por $\neg p$ (y leído como “no **p**”).

La tabla de verdad asociada es por tanto la siguiente:

p	$\neg p$
V	F
F	V

Ejemplo

- Si **p** es “los burros vuelan”, entonces $\neg p$ es “los burros no vuelan”.
- Si **p** es “la puerta está abierta”, entonces $\neg p$ es “la puerta está cerrada”.
- En los números enteros, si **p** es “ser par”, entonces $\neg p$ es “ser impar”.
- En los números reales, si **p** es “ser mayor que 2”, entonces $\neg p$ es “ser menor o igual que 2”.

1.1.2.2. Propiedades y resultados

A continuación, enunciamos una serie de propiedades relativas a los conectores lógicos, resultantes de sus posibles combinaciones. Al darse una clara dualidad con un caso particular de álgebra de Boole sobre las propiedades obtenidas anteriormente para estas son igualmente válidas para este caso, adaptando convenientemente la notación para el caso de la negación, que hace el papel de complementario.

Propiedades

Sean **p, q, r** enunciados lógicos cualesquiera. Entonces, se cumplen las propiedades siguientes:

1. Propiedad asociativa: $p \vee (q \vee r) = (p \vee q) \vee r$, $p (q \wedge r) = (p \wedge q) \wedge r$.
2. Propiedad commutativa: $p \vee q = q \vee p$, $p \vee q = q \wedge p$.
3. Propiedad distributiva: $p \wedge (q \vee r) = (p \wedge q) \vee (p \wedge r)$, $p \vee (q \wedge r) = (p \vee q) \wedge (p \vee r)$.
4. Negación: $p \vee \neg p$ siempre es verdadera, mientras que $p \wedge \neg p$ siempre es falsa.
5. Idempotencia: $p \vee p = p$, $p \wedge p = p$.
6. Absorción: $p \vee (p \wedge q) = p$, $p \wedge (p \vee q) = p$.
7. Propiedad cancelativa:

$$\left. \begin{array}{l} p \vee q \leftrightarrow p \vee r \\ p \wedge q \leftrightarrow p \wedge r \end{array} \right\} \rightarrow q \leftrightarrow r.$$

8. Doble negación: $\neg(\neg p) = p$
9. Leyes de Morgan: $\neg(p \vee q) = \neg p \wedge \neg q$, $\neg(p \wedge q) = \neg p \vee \neg q$.

Asimismo, puede comprobarse que se da el siguiente resultado, que relaciona las implicaciones lógicas con una combinación de disyunciones y negaciones.

Teorema

Sean **p** y **q** dos enunciados lógicos. Entonces el nuevo enunciado lógico **$p \rightarrow q$** puede escribirse de forma equivalente como **$\neg p \vee q$** . Es decir:

$$p \rightarrow q \leftrightarrow \neg p \vee q.$$

El resultado anterior sugiere, por tanto, que todo enunciado lógico complejo puede acabar reduciéndose a otro enunciado contenido únicamente disyunciones, conjunciones y negaciones; incluso, utilizando las leyes de Morgan, solo disyunciones y negaciones (o bien solo conjunciones y negaciones).

Ejemplos:

1. Sea el siguiente enunciado lógico:

$$p \wedge (p \rightarrow q).$$

Vamos a simplificarlo sustituyendo la implicación lógica por la correspondiente expresión equivalente vista en el resultado anterior:

$$p \wedge (\neg p \vee q).$$

Ahora hacemos uso de la propiedad distributiva:

$$(p \wedge \neg q) \vee (p \wedge q).$$

Ahora bien, la proposición lógica **p V \neg q** siempre es falsa, por lo que no tiene efecto en la disyunción anterior, con lo cual la expresión anterior es equivalente a:

$$p \wedge q.$$

2. Tomamos ahora la siguiente proposición lógica:

$$(\neg p \wedge q) \rightarrow (p \vee r).$$

Nuevamente reemplazando la implicación lógica, se tiene:

$$\neg(\neg p \wedge q) \vee (p \vee r).$$

Empleando la ley de Morgan para la disyunción en el primer miembro:

$$(\neg(\neg p) \vee q) \vee (p \vee r).$$

Ahora, por la doble negación, se obtiene:

$$(p \vee q) \vee (p \vee r).$$

Combinando las propiedades conmutativa, distributiva e idempotencia, se tiene:

$$p \vee q \vee r.$$

3. Sea la proposición lógica siguiente:

$$(p \rightarrow q) \rightarrow (p \wedge \neg q).$$

Sustituyendo la implicación lógica exterior, se tiene:

$$\neg(p \rightarrow q) \vee (p \wedge \neg q).$$

Reemplazando ahora la implicación lógica asociada al primer miembro:

$$\neg(\neg p \vee q) \vee (p \wedge \neg q).$$

Utilizando ahora la ley de Morgan en el primer miembro:

$$(\neg(\neg p) \wedge \neg q) \vee (p \wedge \neg q).$$

Haciendo uso de la doble negación:

$$p \wedge \neg q \vee (p \wedge \neg q).$$

Como ambos miembros de la disyunción exterior son idénticos, empleando la idempotencia se tiene:

$$p \wedge \neg q.$$

4. Consideramos el enunciado lógico siguiente:

$$(p \wedge q) \rightarrow (\neg p \rightarrow q).$$

Reemplazando la implicación lógica exterior, tenemos:

$$\neg(p \wedge q) \vee (\neg p \rightarrow q).$$

Aplicando la ley de Morgan en el primer miembro:

$$(\neg p \vee \neg q) \vee (\neg p \rightarrow q).$$

Por otra parte, reemplazando la implicación asociada al segundo miembro:

$$(\neg p \vee \neg q) \vee (\neg(\neg p \vee q)).$$

Empleando de nuevo la doble negación, resulta:

$$(\neg p \vee \neg q) \vee (p \vee q).$$

Combinando adecuadamente los enunciados resultantes mediante la propiedad commutativa y distributiva, se obtiene:

$$(p \vee \neg q) \vee (q \vee \neg q).$$

En este caso, tanto el primer miembro como el segundo miembro de la disyunción exterior son siempre verdaderos por la propiedad de negación, por lo que, al resultar una disyunción entre dos afirmaciones verdaderas, el resultado es también verdadero siempre, obteniéndose por tanto una tautología.

5. Tomamos la proposición lógica siguiente:

$$(p \rightarrow q) \wedge (p \wedge \neg q).$$

Sustituyendo la implicación lógica:

$$(\neg p \vee q) \wedge (p \wedge \neg q).$$

Denotamos por , por lo que tenemos:

$$(\neg p \vee q) \wedge r.$$

Aplicando ahora la propiedad distributiva, se tiene:

$$(\neg p \wedge r) \vee (q \wedge r).$$

Deshaciendo el cambio anterior, , obtenemos:

$$(\neg p \wedge (p \wedge \neg q)) \vee (q \wedge (p \wedge \neg q)).$$

Haciendo uso convenientemente de las propiedades conmutativa, distributiva e idempotencia, resulta:

$$((\neg p \wedge q) \wedge \neg q) \vee (p \wedge (\neg q \wedge q)).$$

No obstante, observamos que tanto el primer miembro de la disyunción como el segundo no se pueden dar, en el primer caso porque p y $\neg q$ no se pueden dar simultáneamente y, en el segundo caso, porque $\neg q$ tampoco se pueden dar simultáneamente. Así pues, en este caso se obtiene una contradicción.

1.2. Álgebra de Boole

El álgebra de Boole es una noción matemática basada en una estructura algebraica que aparece en el año 1847 de la mano del matemático inglés George Boole, en un intento de utilizar las herramientas algebraicas para abordar el estudio de la lógica proposicional. Actualmente, es ampliamente utilizada dentro del campo de diseño electrónico.

1.2.1. Definición y ejemplos

Sean un conjunto B . Se dice B tiene estructura de álgebra de Boole si en B hay definidas dos operaciones internas \vee y \wedge tales que se cumple:

- Para cualquier $x,y,z \in B$, ambas operaciones cumplen la **propiedad asociativa**:

$$\forall x, y, z \in B, x \vee (y \vee z) = (x \vee y) \vee z, x \wedge (y \wedge z) = (x \wedge y) \wedge z.$$

- Para cualquier $x,y,z \in B$, ambas operaciones cumplen la **propiedad conmutativa**:

$$\forall x, y \in B, x \vee y = y \vee x, x \wedge y = y \wedge x.$$

- Para cualquier $x,y,z \in B$, ambas operaciones cumplen la **propiedad distributiva** entre sí:

$$\forall x, y, z \in B, x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z), x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z).$$

- Existen elementos neutros para cada operación, es decir, $\exists 0,1 \in B$ (cero y uno, respectivamente) de forma que $\forall x \in B$ se cumple que $x \vee 0 = x$ y $x \wedge 1 = x$.
- Cada elemento de B tiene asociado un complementario, esto es, $\forall x \in B \exists x' \in B$ tal que $x \vee x' = 1$ y $x \wedge x' = 0$.

Ejemplos:

- Consideremos el conjunto binario $B = \{0,1\}$. Entonces es un álgebra de Boole con las siguientes operaciones \vee y \wedge :

V	0	1
0	0	1
1	1	1

\wedge	0	1
0	0	1
1	1	1

Puede comprobarse que las operaciones anteriormente definidas constituyen un álgebra de Boole, siendo $0 = 1 \vee 1 = 0$.

Por otra parte, este ejemplo constituye las operaciones elementales asociadas a la lógica proposicional, siendo la disyunción (OR) y la conjunción (AND), siendo el complementario la negación lógica.

- Manteniendo la misma notación que en el ejemplo anterior, el conjunto $B^n = \{(x_1, x_2, \dots, x_n) / x_i \in B\}$ es un álgebra de Boole con las operaciones dadas por:

$$(x_1, x_2, \dots, x_n) \vee (y_1, y_2, \dots, y_n) = (x_1 \vee y_1, x_2 \vee y_2, \dots, x_n \vee y_n), \\ (x_1, x_2, \dots, x_n) \wedge (y_1, y_2, \dots, y_n) = (x_1 \wedge y_1, x_2 \wedge y_2, \dots, x_n \wedge y_n).$$

En este caso, se tiene que los elementos cero y uno vienen dados por \neg , respectivamente, siendo el complementario de cualquier elemento.

$$\overline{(x_1, x_2, \dots, x_n)} = (\overline{x_1}, \overline{x_2}, \dots, \overline{x_n}).$$

- Sea A un conjunto cualquiera. Entonces el conjunto de sus partes, $B=P(A)$, con $P(A)=\{C / C \subseteq A\}$, es un álgebra de Boole, tomando $\vee=U$, la unión de conjuntos y $\wedge=\cap$, la intersección de conjuntos. En este caso, cada elemento $C \in P(A)$ tiene asociado un complementario $\bar{C} \in P(A)$, que resulta ser el contrario o conjunto complementario de C con respecto a $P(A)$, es decir, $\bar{C}=\{x \in P(A) / x \notin C\}$. Por último, los correspondientes elementos neutros son $0=\emptyset$ y $1=A$, puesto que $\forall C \in P(A)$ se tiene que $C \cup \emptyset=C$ y $C \cap A=C$, pues por definición $C \subseteq A$.

- Consideremos ahora el conjunto de todos los divisores de 10, que es $B=\{1, 2, 5, 10\}$. Entonces B es un álgebra de Boole con las operaciones $x \vee y = \text{mcm}(x, y)$ y $x \wedge y = \text{mcd}(x, y)$, con $x, y \in B$. En este caso, el elemento cero (neutro para \vee) es 10, mientras que el elemento uno (neutro para \wedge) es . Por otra parte, puede comprobarse que el complementario de un número $x \in B$ es $\bar{x} = \frac{10}{x}$. Así pues, por ejemplo

$$\bar{S} = \frac{10}{S} = 2$$

En general, para cualquier número de la forma $n=p_1 \cdot p_2 \dots p_k$, con p_i primos, $1 \leq i \leq k$, se tiene que el conjunto de sus divisores con las operaciones anteriormente definidas es un álgebra de Boole, con elemento cero 1 y elemento uno n , siendo $\bar{x} = \frac{n}{x}$. Esto se debe a que en este caso el conjunto de sus divisores, con las operaciones establecidas, puede identificarse con el segundo ejemplo, teniendo en cuenta que un divisor cualquiera de n , x , puede escribirse como $n=p_1^{a_1} \cdot p_2^{a_2} \dots p_k^{a_k}$, con $a_i \in \mathbb{B}=\{0,1\}$, $1 \leq i \leq k$.

Por el contrario, si en la descomposición aparece algún primo con una potencia superior a 1, entonces el resultado no es cierto. Por ejemplo, tomamos $n=12=2^2 \cdot 3$, cuyos divisores son $\{1, 2, 3, 6, 12\}$. Entonces, si cogemos por ejemplo $x=2$, tenemos que x no tiene complementario, puesto que no existe ningún elemento \bar{x} tal que $x \wedge \bar{x} = \text{mcd}(2, \bar{x}) = 1$ y a la vez $x \vee \bar{x} = \text{mcm}(2, \bar{x}) = 12$.

1.2.2. Propiedades y resultados

Propiedades

Sea (B, V, Λ) un álgebra de Boole y $x, y, z \in B$. Entonces se cumplen las siguientes propiedades:

1. Idempotencia: $x \vee x = x$, $x \wedge x = x$.
2. Dominación: $x \vee 1 = x$, $x \wedge 0 = x$.
3. Absorción: $x \vee (x \wedge y) = x$, $x \wedge (x \vee y) = x$.
4. Propiedad cancelativa:

$$\left. \begin{array}{l} x \vee y = x \vee z \\ x \wedge y = x \wedge z \end{array} \right\} \rightarrow y = z.$$

5. Doble complementario: $\bar{\bar{x}} = x$.
6. Leyes de Morgan: $\overline{x \vee y} = \bar{x} \wedge \bar{y}$, $\overline{x \wedge y} = \bar{x} \vee \bar{y}$.

Notas

1. La cuarta propiedad garantiza la existencia de un único elemento complementario; es decir, se puede afirmar que si B es un álgebra de Boole, entonces $\forall x \in B \exists ! \bar{x} \in B / x \vee \bar{x} = 1$ y $x \wedge \bar{x} = 0$.
2. La elección de los símbolos V y Λ que se han utilizado como operaciones sobre álgebras de Boole se debe a la búsqueda de una consonancia con la notación que utilizaremos en la lógica proposicional, aunque también suelen utilizarse los símbolos suma (+) y producto (\cdot), respectivamente.

El siguiente resultado está en consonancia con el segundo ejemplo presentado anteriormente, y permite generar nuevas álgebras de Boole a partir de dos o más álgebras de Boole.

Teorema

Sean (B_1, V_1, Λ_1) y (B_2, V_2, Λ_2) dos álgebras de Boole. Entonces, el conjunto dado por $B = B_1 \times B_2$ tiene estructura de álgebra de Boole con las siguientes operaciones:

$$(x, y) \vee (x', y') = (x V_1 x', y V_2 y'), (x, y) \wedge (x', y') = (x \Lambda_1 x', y \Lambda_2 y').$$

El resultado anterior puede generalizarse fácilmente para n álgebras de Boole.

1.3. Funciones booleanas y puertas lógicas

En esta sección vamos a definir y trabajar con las funciones booleanas, para posteriormente pasar a definir a partir de ellas las puertas lógicas, con las cuales se pueden construir circuitos lógicos. Asimismo, por comodidad en esta sección emplearemos una notación diferente para hacer referencia a las operaciones, reemplazándola por la indicada en la segunda nota de la sección anterior. Es decir, a partir de ahora reemplazaremos el símbolo \vee por \wedge y el símbolo $+$ por \cdot . De esa forma, el álgebra de Boole asociada al conjunto \mathbb{B} tiene las siguientes operaciones:

$+$	0	1
0	0	1
1	1	1

\cdot	0	1
0	0	1
1	1	1

1.3.1. Funciones booleanas

Una función booleana con n variables es una aplicación $f:\mathbb{B}^n \rightarrow \mathbb{B}$. Se denota por F_n al conjunto formado por todas las funciones booleanas de n variables, es decir:

$$F_n = \{f:\mathbb{B}^n \rightarrow \mathbb{B} / f \text{ aplicación}\}$$

Ejemplos:

1. Sea la aplicación $f:\mathbb{B}^n \rightarrow \mathbb{B}$ dada por $f(0)=0$ y $f(1)=1$. Esta aplicación responde a la expresión $f(x)=x$, ya que lleva a todo a sí mismo, por lo que se trata de la aplicación identidad.
2. Consideremos ahora la aplicación $f:\mathbb{B}^n \rightarrow \mathbb{B}$ definida por $f(0)=1$ y $f(1)=0$. En esta ocasión, la aplicación responde a la expresión $f(x)=\bar{x}$, es decir, es aquella que lleva un elemento de \mathbb{B} a su correspondiente elemento complementario.
3. Sea $f:\mathbb{B}^2 \rightarrow \mathbb{B}$ dada por $f(x,y)=x+y$. En este caso, f actúa como la operación \vee y se tienen, por tanto, las siguientes identidades:
 - a. $f(0,0)=0+0=0$.
 - b. $f(0,1)=0+1=1$.
 - c. $f(1,0)=1+0=1$.
 - d. $f(1,1)=1+1=1$.
4. Sea $f:\mathbb{B}^2 \rightarrow \mathbb{B}$ dada por $f(x,y)=x \cdot y$. Esta vez, f actúa como la operación \wedge , donde se tienen las siguientes asignaciones:
 - a. $f(0,0)=0 \cdot 0=0$.
 - b. $f(0,1)=0 \cdot 1=0$.
 - c. $f(1,0)=1 \cdot 0=0$.
 - d. $f(1,1)=1 \cdot 1=1$.

Obsérvese que en los dos últimos ejemplos anteriores se ha podido identificar cada una de las dos funciones booleanas de dos variables con cada una de las dos operaciones definidas anteriormente. En general, las funciones booleanas con dos variables pueden utilizarse para definir a partir de ellas nuevas operaciones que resultan de utilidad.

Asimismo, las funciones booleanas por sí solas representan una nueva estructura a la cual se le puede asignar un álgebra de Boole.

Por otro lado, sean $f, g: \mathbb{B}^n \rightarrow \mathbb{B}$ funciones booleanas; o lo que es lo mismo, $f, g \in F_n$. Entonces definimos las siguientes operaciones sobre F_n :

- Por una parte, se define la **suma** de funciones booleanas con n variables, como una nueva función booleana, $f+g \in F_n$, dada por:

$$(f+g)(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n) + g(x_1, x_2, \dots, x_n).$$

- Por otra parte, podemos definir el **producto** de funciones booleanas con n variables, $f \cdot g$, como una nueva función booleana, $f \cdot g \in F_n$, dada por:

$$(f \cdot g)(x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n) \cdot g(x_1, x_2, \dots, x_n).$$

Con estas operaciones definidas sobre F_n , puede comprobarse que se obtiene una estructura de álgebra de Boole, tal y como se afirma en el resultado siguiente.

Teorema

La terna $(F_n, +, \cdot)$, siendo $+$ y \cdot las operaciones definidas anteriormente, es un álgebra de Boole, siendo las funciones constantes cero y uno sus respectivos elementos neutros y el complementario de una función $f \in F_n$, la función $\bar{f} \in F_n$ dada por:

$$\bar{f}(x_1, x_2, \dots, x_n) = \overline{f(x_1, x_2, \dots, x_n)}.$$

Ejemplo:

$(F_1, +, \cdot)$ es un álgebra de Boole que consta de cuatro elementos, desglosados a continuación:

- Aplicación identidad: $Id: \mathbb{B} \rightarrow \mathbb{B}$ dada por $Id(x)=x$. Se tiene, por tanto:
 - $Id(0)=0$,
 - $Id(1)=1$.
- Elemento cero: $f: \mathbb{B} \rightarrow \mathbb{B}$ dada por $f(0)=0$. Se tiene, entonces:
 - $f(0)=0$,
 - $f(1)=0$.
- Elemento uno: $g: \mathbb{B} \rightarrow \mathbb{B}$ definida por $g(1)=1$. Por tanto, se tiene:
 - $g(0)=1$,
 - $g(1)=1$.

- Aplicación complementaria: $h:\mathbb{B} \rightarrow \mathbb{B}$ dada por $h(x)=\bar{x}$. Así pues, se tiene:
 - $h(0)=1$,
 - $h(1)=0$.

Se obtiene la siguiente tabla de operaciones.

$+$	Id	f	g	h
Id	Id	Id	g	g
f	Id	f	g	h
g	g	g	g	g
h	g	h	g	h

\cdot	Id	f	g	h
Id	Id	f	Id	f
f	f	f	f	f
g	Id	f	f	h
h	f	f	f	h

Bien sea a partir de la definición, o bien a partir de la observación de ambas tablas, pueden deducirse los complementarios de cada elemento:

- $\bar{Id}=h$.
- $\bar{f}=g$.
- $\bar{g}=f$.
- $\bar{h}=Id$.

1.3.2. Puertas lógicas

Las puertas lógicas consisten en una serie de operadores lógicos tales que, dados un par de valores binarios, es decir, $x, y \in \mathbb{B}$, estos devuelven un valor binario, que será 0 o 1 en función del valor de x e y . Asimismo, también existen algunos que actúan solo sobre un único elemento $x \in \mathbb{B}$. Por tanto, las puertas lógicas pueden considerarse unos operadores binarios, que se obtienen como caso particular de las funciones booleanas con dos variables, o lo que es lo mismo, como un subconjunto de F_2 , para el caso en el que operen sobre dos valores binarios, mientras que, para el caso de una variable, pueden verse como un subconjunto de F_1 .

Veamos a continuación las puertas lógicas más relevantes, las cuales constituyen los fundamentos de la computación, por tratarse de las encargadas de realizar las operaciones aritméticas a partir de información binaria, asociadas con el paso de corriente eléctrica (1) o impedimento de este (0) en función de la señal de entrada (es decir, los valores de x y/o y).

- **Puerta OR**

Se trata de una puerta lógica que envía un par de valores $x, y \in \mathbb{B}$ a otro valor de \mathbb{B} , respondiente a la tabla de operaciones que se muestra a continuación:

OR	$y = 0$	$y = 1$
$x = 0$	0	1
$x = 1$	1	1

La tabla anterior ilustra que esta puerta lógica puede identificarse con el operador $+$ del álgebra de Boole sobre \mathbb{B} , trabajada en esta sección. Asimismo, se observa que para que el resultado de esta operación sea 1 basta con que alguna de las dos variables (o ambas) sea 1; dicho de otra forma, el paso de corriente se permite siempre y cuando pase corriente por alguno de los dos parámetros.

Por una parte, este operador puede interpretarse como la función booleana definida de F_2 , relacionada con la operación $,$ en términos de la identificación anteriormente establecida:

$$f: \mathbb{B}^2 \rightarrow \mathbb{B}, f(x, y) = x + y.$$

Por otra parte, el operador OR puede esquematizarse como un circuito eléctrico con dos interruptores dispuestos en paralelo, de forma que para lograr el paso de corriente basta con que uno de los interruptores esté encendido (o ambos simultáneamente).

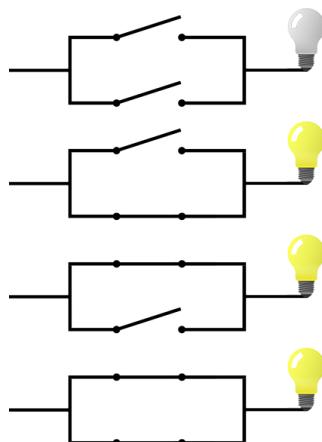


Figura 1. Fragmento de circuito eléctrico en representación de la puerta lógica OR. Fuente: elaboración propia.

Por último, la puerta lógica OR se representa mediante el símbolo que se muestra a continuación:

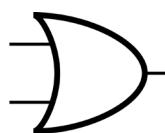


Figura 2. Símbolo de la puerta lógica OR. Fuente: elaboración propia.

- **Puerta AND**

Esta puerta lógica envía un par de valores $x, y \in \mathbb{B}$ a otro valor de \mathbb{B} , siguiendo la tabla de operaciones siguiente:

AND	$y = 0$	$y = 1$
$x = 0$	0	0
$x = 1$	0	1

Obsérvese que esta puerta lógica coincide con el operador del álgebra de Boole sobre \mathbb{B} con la que ya trabajamos. Además, en este caso puede apreciarse que el resultado únicamente es 1 cuando ambas variables lo son; es decir, la salida de corriente solo está permitida cuando por ambos parámetros de entrada pasa corriente.

Por otra parte, este operador puede identificarse con la siguiente función booleana en F_2 , asociada con su relación con la operación producto, que anteriormente se ha mencionado:

$$f: \mathbb{B}^2 \rightarrow \mathbb{B}, f(x, y) = x \cdot y.$$

Asimismo, el operador AND puede esquematizarse como un circuito eléctrico con dos interruptores dispuestos en serie, de forma que la única forma para la cual puede pasar la corriente es si ambos interruptores están encendidos.

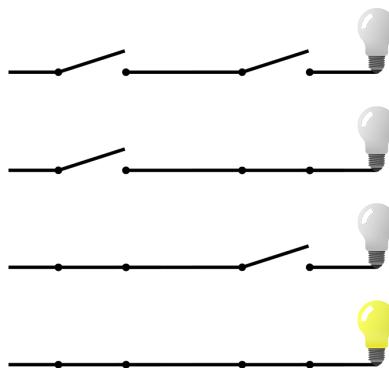


Figura 3. Fragmento de circuito eléctrico en representación de la puerta lógica AND. Fuente: elaboración propia.

Para terminar, la puerta lógica AND viene representada a través del símbolo siguiente:

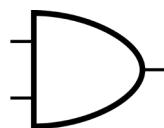


Figura 4. Símbolo de la puerta lógica AND. Fuente: elaboración propia.

• Puerta XOR

Recibe comúnmente el nombre de **exclusive OR** o, lo que es lo mismo, de “o exclusivo”. Se diferencia de la puerta OR anteriormente expuesta por el hecho de que, como su propio nombre indica, excluye el caso en el que ambas variables sean 1, permitiendo de ese modo el paso únicamente cuando solo alguno de los dos valores es 1. Por tanto, la tabla de operaciones es la siguiente:

XOR	$y = 0$	$y = 1$
$x = 0$	0	1
$x = 1$	1	0

Puede comprobarse que este operador puede representarse como la siguiente función booleana en F_2 :

$$f: B^2 \rightarrow B, f(x, y) = x \oplus y = \overline{x} \cdot y + x \cdot \overline{y} \equiv (\overline{x} + y) \cdot (x + \overline{y}).$$

La puerta lógica XOR suele representarse a través del siguiente símbolo:

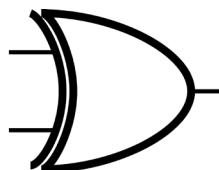


Figura 5. Símbolo de la puerta lógica XOR. Fuente: elaboración propia.

- **Puerta NOT**

Esta puerta lógica actúa sobre una única variable de entrada y la transforma en su valor contrario (o complementario); es decir, dado, $x \in \mathbb{B}$ esta puerta lógica lo envía al valor $\bar{x} \in \mathbb{B}$, por lo que puede identificarse con la siguiente función booleana en :

$$f: B \rightarrow B, f(x) = \bar{x}.$$

Por tanto, su tabla operacional es la siguiente:

NOT	
$x = 0$	1
$x = 1$	0

La puerta lógica NOT también tiene representación a través del símbolo que se muestra a continuación:

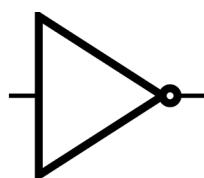


Figura 6. Símbolo de la puerta lógica NOT. Fuente: elaboración propia.

- **Puerta NOR**

Consiste en una puerta lógica que envía un par de valores $x, y \in \mathbb{B}$ a otro valor de , consistente en invertir el resultado de salida de la puerta lógica OR, vista con anterioridad. Por tanto, presenta la tabla de operaciones siguiente:

NOR	$y = 0$	$y = 1$
$x = 0$	1	0
$x = 1$	0	0

En este caso, la puerta únicamente genera corriente (es decir, obtiene como resultado 1) en el caso en el que ninguno de los dos parámetros lleve corriente, o lo que es lo mismo, en caso de que ambos parámetros de entrada sean 0.

Por otra parte, puede comprobarse que este operador puede representarse como la siguiente función booleana de F_2 :

$$f: \overline{B}^2 \rightarrow B, f(x, y) = x + y.$$

La puerta lógica NOR se representa habitualmente mediante el símbolo que se muestra a continuación:

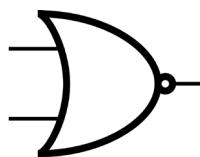


Figura 7. Símbolo de la puerta lógica NOR. Fuente: elaboración propia.

- **Puerta NAND**

Se trata de una puerta lógica tal que, dado un par de valores de entrada $x, y \in \overline{B}$, produce como resultado otro valor de \overline{B} , el cual consiste en una inversión del resultado de salida producido por la puerta lógica AND, vista anteriormente. Así pues, tiene la tabla de operaciones que se muestra a continuación:

AND	$y = 0$	$y = 1$
$x = 0$	1	1
$x = 1$	1	0

De la tabla de valores anterior puede deducirse que esta puerta lógica genera corriente siempre y cuando no haya paso de corriente en ambos valores de entrada; es decir, siempre generará como resultado 1 salvo cuando ambos parámetros de entrada sean 1.

Asimismo, se puede comprobar que este operador tiene asociada la siguiente función booleana de F_2 :

$$f: \overline{B}^2 \rightarrow B, f(x, y) = x \cdot y.$$

La puerta lógica NAND suele representarse a través del símbolo que se muestra a continuación:

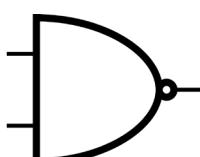


Figura 8. Símbolo de la puerta lógica NAND. Fuente: elaboración propia.

- **Puerta XNOR**

Consiste en una puerta lógica en la que, dado un par de valores de entrada $x, y \in \mathbb{B}$, produce otro valor en \mathbb{B} consistente en la inversión del resultado de salida producido por la puerta lógica XOR, ya expuesta con anterioridad. Por tanto, su tabla de operaciones asociada es la siguiente:

XNOR	y = 0	y = 1
x = 0	1	01
x = 1	0	1

Tal y como puede comprobarse en la tabla, esta puerta lógica únicamente permite el paso de corriente en el caso en el que ambos parámetros de entrada estén en el mismo estado, es decir, los dos sin corriente o bien ambos con corriente. Es decir, su parámetro de salida es 1 en caso de que los dos parámetros de entrada sean 0 o bien ambos sean 1.

Por otra parte, se tiene que este operador tiene asociada la función booleana de F_2 que se presenta a continuación:

$$\begin{aligned} f : B^2 \rightarrow B, f(x) &= (x \oplus y) = (\bar{x} + \bar{y}) \cdot (x + y) = (\bar{x} + \bar{y}) + (x + y) = \bar{x} \cdot \bar{y} + \bar{x} \cdot y \\ &= x \cdot y + x \cdot y. \end{aligned}$$

La puerta lógica XNOR suele representarse mediante el símbolo que se presenta a continuación:

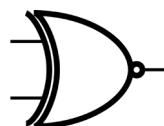


Figura 9. Símbolo de la puerta lógica XNOR. Fuente: elaboración propia.

1.3.3. Circuitos lógicos

Todas estas puertas lógicas pueden combinarse entre sí para construir circuitos lógicos, de forma que, en última instancia, constituyen los fundamentos más básicos de la computación, a partir de los cuales se construyen las estructuras que permiten la realización de operaciones aritméticas, presentes en los dispositivos electrónicos más rudimentarios, como calculadoras básicas, como en aquellos de mayor complejidad, como ordenadores.

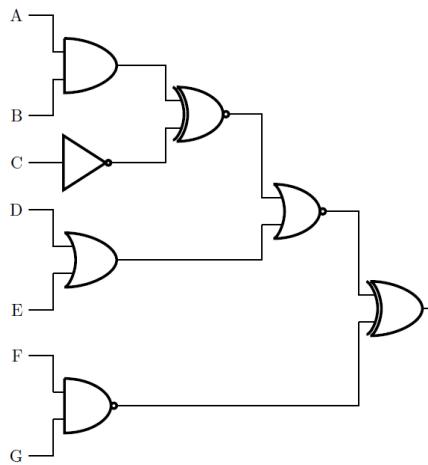


Figura 10. Ejemplo de circuito lógico. Fuente: elaboración propia.

La naturaleza de los circuitos lógicos, basada en la combinación de una serie de puertas lógicas, puede verse como una función determinada que envía un conjunto de parámetros booleanos determinados (señal de entrada) y produce como salida otro conjunto de valores booleanos (señal de salida). Veamos algunos ejemplos en los que se ilustra cómo realizar el proceso de interpretación de la forma en que se manipula la información en dichos circuitos.

Ejemplos:

Vamos a calcular la función booleana correspondiente a la salida de una serie de circuitos que mostramos a continuación. Para ello, calcularemos las expresiones algebraicas que resultan de estos.

1. Consideremos el circuito mostrado en la imagen siguiente:

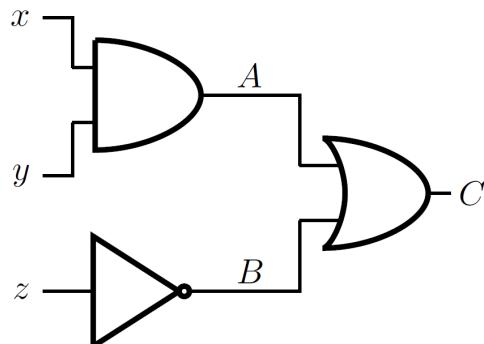


Figura 11. Circuito con puertas lógicas AND (A), NOT (B) y OR (C). Fuente: elaboración propia.

Entonces las expresiones algebraicas correspondientes a cada tramo del circuito son las siguientes:

- En lo que respecta a A, se tiene que es el resultado de pasar x e y por una puerta AND, por lo que se tiene que $A = x \cdot y$. Por otra parte, es el resultado de pasar z por una puerta NOT, luego $B = \bar{z}$.
- En cuanto a C, se trata del resultado tras hacer pasar A y B por una puerta OR, por lo que $C = A + B$. Ahora bien, como $A = x \cdot y$ y $B = \bar{z}$, se tiene que $C = x \cdot y + \bar{z}$, luego la función booleana asociada a este circuito es $f \in F_3$ dada por:

$$f(x, y, z) = x \cdot y + \bar{z}.$$

2. Tomamos ahora el circuito que se muestra en la ilustración siguiente:

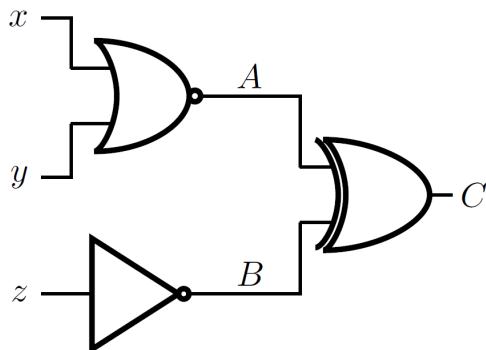


Figura 12. Circuito con puertas lógicas NOR (A), NOT (B) y XOR (C).Fuente: elaboración propia.

Se tiene que en este caso la expresión algebraica correspondiente al circuito considerado se calcula como sigue:

- 1) En primer lugar, se tiene que A es el resultado de pasar por la puerta lógica NOR los parámetros x e y, luego, utilizando la correspondiente ley de Morgan, $A = \bar{x} + \bar{y} = \bar{x} \cdot \bar{y}$. Asimismo B, consiste en el resultado de pasar el parámetro z por una puerta NOT, por lo que $B = \bar{z}$.
- 2) Por último, en lo que a C se refiere, se tiene que su salida es el resultado de pasar A y B por una puerta XOR, con lo cual $C = A \oplus B$. Ahora, teniendo en cuenta que $A = \bar{x} \cdot \bar{y}$ y que $B = \bar{z}$, tenemos que $C = (\bar{x} \cdot \bar{y}) \oplus \bar{z}$, con lo que haciendo uso de la definición se obtiene $C = ((\bar{x} \cdot \bar{y}) + \bar{z}) \cdot (\bar{x} \cdot \bar{y} + \bar{z})$. Utilizando nuevamente la ley de Morgan correspondiente y la cancelación de la doble negación, se tiene que $C = (x + y + z) \cdot (\bar{x} \cdot \bar{y} + \bar{z})$. Así pues, este circuito lógico puede representarse a través de la función $f \in F_3$ definida como:

$$f(x, y, z) = (x + y + z) \cdot (\bar{x} \cdot \bar{y} + \bar{z}).$$

3. Sea el circuito lógico mostrado en la figura siguiente:

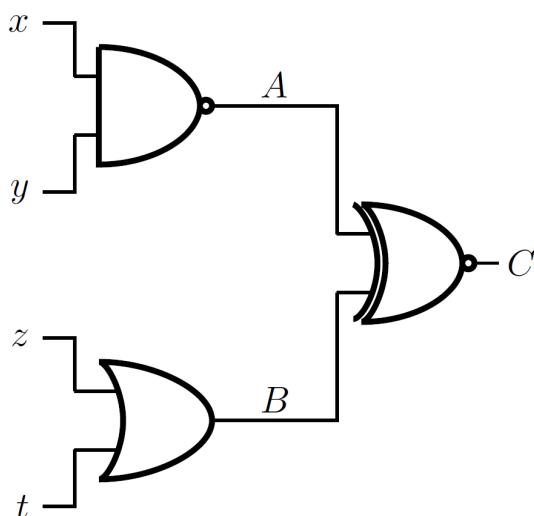


Figura 13. Circuito con puertas lógicas NAND (A), OR (B) y XNOR (C).Fuente: elaboración propia.

En este caso, la expresión algebraica asociada al circuito considerado se calcula de la siguiente forma:

- 1) Por una parte, A corresponde al resultado pasar por la puerta lógica NAND los parámetros x e y, luego se tiene $A = (x \cdot y) = \bar{x} + \bar{y}$, donde en la última igualdad se ha hecho uso de la ley de Morgan correspondiente. A su vez, B es el resultado de pasar por la puerta lógica OR z y t, con lo cual $z + t$.
- 2) Por otra parte, C resulta de pasar por la puerta lógica XNOR los parámetros A y B obtenidos anteriormente, por lo que $C = \bar{A} \oplus \bar{B} = A \cdot B + \bar{A} \cdot \bar{B}$. Como $A = \bar{x} + \bar{y}$ y $B = z + t$, se tiene que $C = (\bar{x} + \bar{y}) \cdot (z + t) + (\bar{x} + \bar{y}) \cdot (\bar{z} + \bar{t})$. Ahora, aplicando la ley de Morgan procedente, junto con la cancelación de la doble negación, obtenemos $C = (\bar{x} + \bar{y}) \cdot (z + t) + (\bar{x} + \bar{y}) \cdot (\bar{z} + \bar{t})$. Así pues, este circuito lógico puede identificarse con la función booleana $f \in F_4$ dada por:

$$f(x, y, z, t) = (\bar{x} + \bar{y}) \cdot (z + t) + x \cdot y \cdot \bar{z} \cdot \bar{t}$$

Para profundizar más sobre cuestiones relativas a este tema, puede consultarse la obra de Jiménez, *Matemáticas para la computación*.

1.4. Lógica de predicados

La lógica proposicional constituye una teoría que permite formalizar y estructurar el pensamiento lógico, con aplicación directa a las diferentes disciplinas científicas donde se requiere el uso del razonamiento lógico, especialmente en el ámbito de las matemáticas. No obstante, esta teoría está limitada, en tanto que no es capaz de cubrir la totalidad de situaciones que pueden resolverse mediante el uso de las diferentes técnicas de razonamiento. Veamos algunos ejemplos:

- Afirmaciones como “ $x=2$ ” o “ $x<4$ ” (sin concretar el valor de x) no son proposiciones lógicas como tal, en tanto que su valor de verdad depende del valor que tenga la variable x. Por ejemplo, para $x=2$ ambas afirmaciones son verdaderas, para $x=3$ la primera es falsa y la segunda es verdadera y para $x=4$ ambas son falsas.
- Existen otras afirmaciones complejas, formuladas en lenguaje habitual, que no pueden ser descritas mediante la lógica proposicional. Por ejemplo, consideremos el siguiente par de afirmaciones:
 - “Todas las mascotas son animales”.
 - “Gala es la mascota de Laura”.

Con su correspondiente conclusión: “Por tanto, Gala es un animal”.

No obstante, el argumento que nos ha permitido llegar a tal conclusión no se puede representar mediante los elementos que nos proporciona la lógica proposicional.

En conclusión, es necesario recurrir a otras herramientas que nos permitan trabajar con mayor generalidad ante situaciones como las expuestas anteriormente. Si observamos con atención el segundo caso, observamos que la dificultad estriba en la existencia de **predicados**, como acompañantes del sujeto sobre el cual se realiza la acción (o bien se describe). Así pues, en este apartado nos encargaremos de ilustrar un nuevo tipo de lógica dedicada al estudio de este tipo de afirmaciones, que, por lo anteriormente mencionado, recibe el nombre de **lógica de predicados** o **lógica de primer orden**, la cual se encargará de analizar la interacción entre el sujeto y el predicado sujetos a estudio.

1.4.1. Predicados y universos

Un predicado $p(x_1, x_2, \dots, x_n)$ es una afirmación que hace referencia a una propiedad o una relación entre objetos x_1, x_2, \dots, x_n . Dichas afirmaciones disponen de un valor de verdad al reemplazar las variables (objetos) por valores específicos, c_1, c_2, \dots, c_n , de forma que la afirmación resultante de reemplazar el predicado por los correspondientes valores considerados, $p(c_1, c_2, \dots, c_n)$, puede ser verdadera o falsa.



Ejemplo

1. La afirmación $p(x): "x \text{ es un reptil}"$ consiste en un predicado, cuyo valor de verdad está sujeto al valor que se le asigne a x . Así pues, si, por ejemplo, reemplazamos x por “serpiente” el enunciado es verdadero, mientras que si lo hacemos por “gato” el predicado pasa a ser falso.
2. Los predicados se utilizan constantemente en los lenguajes de programación de alto nivel. Por ejemplo, considérese el enunciado “si, $x > 0$ definir $y=1$; en caso contrario, definir $y=0$ ”, que contiene el predicado $p(x): "x > 0"$. Se observa que, por ejemplo, haciendo $x=1$, se tiene que $p(x)$ es verdadero, de modo que ejecuta la instrucción $y=1$, mientras que para $x=-2$, resulta que $p(x)$ es falso, resultando, por tanto, en la asignación $y=0$.
3. En el predicado en cuestión también pueden intervenir más de un objeto; por ejemplo, considérese la afirmación $p(x,y): "x \text{ es padre de } y"$. Si reemplazamos x por Juan e y por María, el enunciado será verdadero siempre y cuando María sea hija de Juan, mientras que será falso en caso contrario.
4. Consideremos el predicado formado por dos objetos $p(x,y): "x+y=3"$. Se tiene, por ejemplo, $p(1, 2)$ es verdadero, ya que $1+2=3$; asimismo, tiene otros valores verdaderos, como $p(2, 1)$, $p(0, 3)$, $p(-2, 5)$, etc. (de hecho tiene un número infinito de ellos). Por otra parte, también tiene valores en los que es falso, como, por ejemplo, $p(1, 1)$, ya que $1+1\neq 3$; de hecho también tiene un número infinito de valores donde no se cumple la condición, como, por ejemplo, $p(2, 4)$, $p(5, -1)$, $p(3, 1)$, etc.

Además del concepto de predicado, conviene tener en cuenta sobre qué rango de valores pueden moverse las variables que intervienen en estos.

El **conjunto universal** o **universo** U es el conjunto (no vacío) de los diferentes valores que pueden tomar las variables que intervienen en la evaluación de un predicado.



Ejemplo

1. Si consideramos el predicado $p(x): "x \text{ es estudiante}"$, tiene sentido tomar U como el conjunto de todos los seres humanos, mientras que no lo tendría tomar otras opciones, como, por ejemplo, el conjunto de todos los colores o el conjunto formado por todos los países del mundo.
2. Por otra parte, si ahora tomamos la afirmación $p(x): "x \text{ es par}"$, tiene sentido tomar como conjunto alguno formado por números, como, por ejemplo, $U=\mathbb{N}$ o $U=\mathbb{Z}$, entre otras posibilidades.

La lógica de predicados permite describir procesos y algoritmos que pueden ejecutarse sin ningún tipo de ambigüedad en cada uno de sus pasos, obteniendo un resultado único. A continuación, se presenta un ejemplo.

Tomemos como universo $U=\mathbb{Z}$ y supongamos que tenemos $x, y \in U$ tales que inicialmente $x=0$ e $y=1$. Describimos a continuación el siguiente algoritmo.

- a. Si $x=1$ o $y=1$, entonces asignar $x=y$ y luego $y=0$.
- b. Si $x=y$, entonces asignar $x=x \cdot y$; de lo contrario, asignar $y=x \cdot y$.
- c. Asignar $x=x+2$ y luego $y=x-y+1$.
- d. Si $x=2$ e $y=4$, entonces asignar $y=x$.

Veamos a continuación la traza del algoritmo, con el fin de comprobar su resultado final.

- a. La afirmación “ $x=1$ o $y=1$ ” es verdadera, ya que aunque la primera igualdad es falsa, la segunda es verdadera, y como se trata de una disyunción (o), basta con que una de las dos afirmaciones sea verdadera para que el resultado lo sea. Así pues, se ejecuta la consecuencia que es asignar $x=y$, por lo que $x=1$, ya que actualmente tiene el valor 1 y, por otra parte, $y=0$. Así pues, al terminar este paso se tiene que $x=1$ e $y=0$.
- b. La afirmación $x=y$ es falsa, puesto que $x=1$ e $y=0$. Por tanto, saltamos a la instrucción $y=x \cdot y$, y, por tanto, $y=1 \cdot 0=0$. En consecuencia, actualmente tenemos $x=1$ e $y=0$.
- c. En este caso no hay ninguna condición, por lo que se ejecutan directamente las instrucciones señaladas: en primer lugar se asigna a x el valor $x+2$, por tanto, $x=1+2=3$ y, por otra parte, se asigna a y el valor $y=x-y+1$, con lo cual $y=3-0+1=4$. Por tanto, al finalizar este paso se tiene que $x=3$ e $y=4$.
- d. Por último, la condición “ $x=2$ e $y=5$ ” no se cumple puesto que, aunque la primera sea cierta, la segunda es falsa, y al tratarse de una conjunción ambas afirmaciones tienen que ser ciertas para que el resultado sea verdadero y, por tanto, la instrucción no se ejecuta. En consecuencia, los valores finales de x e y son los del paso anterior, es decir, $x=2$ e $y=4$.

1.4.2. Cuantificadores

Los cuantificadores son otros de los elementos imprescindibles dentro del ámbito de la lógica de primer orden. Vamos a definir a continuación los dos tipos de cuantificadores, ampliamente usados en el campo de la lógica y en todas y cada una de las ramas de la matemática.

Si $p(x)$ es un predicado con variable x , la afirmación “para todo x , $p(x)$ ” es una proposición que indica que para cualquier valor de $x \in U$, se cumple $p(x)$. En este caso, se dice que la variable x está **universalmente cuantificada**. El símbolo con el cual se denota esta relación es el **cuantificador universal**, denotado por el símbolo \forall , por lo que la afirmación anterior puede escribirse como:

$$\forall x, p(x).$$

En general, puede extenderse a múltiples variables. Por ejemplo, para expresar que para todo x e y se cumple $p(x,y)$, puede denotarse como:

$$\forall x, \forall y, p(x, y)$$

o bien

$$\forall x, y, p(x, y).$$

El valor de verdad de la afirmación $\forall x, p(x)$ se determina a través del siguiente criterio:

1. Si $p(x)$ es una proposición verdadera para todos los valores x del conjunto universal considerado, U , entonces dicha afirmación es verdadera.
2. En caso contrario, si existe al menos un elemento x del conjunto universal U que no cumple $p(x)$, es decir, para el cual la proposición $p(x)$ es falsa, entonces el enunciado es falso.



Ejemplo

Se presenta a continuación una serie de ejemplos en los que una afirmación con predicado puede transcribirse formalmente empleando cuantificadores universales.

1. “Todo conjunto tiene incluido el conjunto vacío en él”:

$$\forall A, \emptyset \subseteq A.$$

La afirmación anterior puede transcribirse como:

$$\forall A, p(A),$$

donde $p(A)$ es el predicado " $\emptyset \subseteq A$ ", es decir, “el vacío está incluido en A ”.

2. “Todo número entero es también un número natural”:

$$\forall x \in \mathbb{Z}, x \in \mathbb{N}.$$

Esta afirmación puede transcribirse en forma general como:

$$\forall x, p(x),$$

donde $U = \{x \mid p(x)\}$ es el predicado “ $x \in \mathbb{N}$ ”, es decir, “ x pertenece al conjunto de los números naturales” o, lo que es lo mismo, “ x es un número natural”.

Puede observarse que el primer ejemplo constituye una afirmación verdadera, puesto que, en efecto, todo conjunto tiene el conjunto vacío como subconjunto de este. Por el contrario, la segunda afirmación es falsa, ya que no todo número entero es un número natural; por ejemplo, $-1 \in \mathbb{Z}$, pero $-1 \notin \mathbb{N}$ (de hecho, hay un número infinito de ejemplos más, tantos como números negativos, pero es suficiente con un ejemplo para demostrar que la afirmación en cuestión es falsa).

Si $p(x)$ es un predicado cuya variable es x , el enunciado “existe un x tal que $p(x)$ ” es una proposición que indica que existe algún valor de U tal que se cumple $p(x)$. En este caso, decimos que x está **existencialmente cuantificada**. El símbolo con el cual se denota este cuantificador es \exists , por lo que la afirmación anterior puede reescribirse de la siguiente forma:

$$\exists x: p(x),$$

o bien

$$\exists x / p(x).$$

Al igual que en el caso anterior, su uso puede extenderse a múltiples variables. Así pues, por ejemplo, para demostrar que existen x e y tales que se cumple $p(x,y)$, escribimos:

$$\exists x, \exists y : p(x,y)$$

o bien

$$\exists x, y : p(x,y).$$

El valor de verdad de la afirmación $\exists x: p(x)$ se determina empleando el siguiente criterio:

1. Si existe algún valor x del conjunto universal U tal que se cumple $p(x)$, entonces la proposición anterior es verdadera.
2. Si, por el contrario, no hay ningún valor de x en U para el cual se cumple $p(x)$, entonces la proposición anterior es falsa.

Veamos seguidamente un par de ejemplos en los que se ilustran diferentes contextos en los que puede utilizarse el cuantificador existencial.

1. “Existe algún número entero negativo”:

$$\exists x \in Z : x < 0.$$

La afirmación anterior puede reescribirse como:

$$\exists x : p(x),$$

donde $U = Z$ y $p(x)$ es el predicado “ $x < 0$ ”, es decir, “ser negativo”.

2. “Existe algún número real tal que al multiplicarlo por 0 el resultado es 1”:

$$\exists x \in R : x \cdot 0 = 1.$$

El enunciado anterior puede transcribirse como:

$$\exists x : p(x),$$

donde $U = R$ y $p(x)$ es el predicado “ $x \cdot 0 = 1$ ”.

Se puede ver que el primer ejemplo consta de una afirmación verdadera, ya que, por ejemplo, $-3 \in Z$ es un valor que cumple $-3 < 0$; de hecho, existe un número infinito más (tantos como enteros negativos hay), aunque basta con encontrar un único caso en el que se cumpla $p(x)$ para probar que la afirmación es verdadera). Por otra parte, el segundo ejemplo constituye una afirmación falsa, ya que no hay ningún número real que multiplicado por 0 dé 1, puesto que, de hecho, $\forall x \in R, x \cdot 0 = 0 \neq 1$.

1.4.3. Semántica y propiedades

Antes de proceder al estudio de enunciados completos, conviene introducir notación y establecer las propiedades y reglas más elementales. En estos términos, las fórmulas de la lógica de primer orden se construyen empleando los siguientes elementos para conectar todos los elementos que intervienen (variables y predicados):

- Conectores lógicos: $\vee, \wedge, \rightarrow, \leftrightarrow, \neg$.
- Paréntesis o corchetes.
- Relación binaria: $=$.
- Variables.
- Cuantificadores: \forall y \exists .

En lo que respecta a las propiedades, conviene reseñar que, al trasformar los predicados en proposiciones sustituyendo las variables por constantes, estas están sujetas a las reglas y propiedades relativas a los conectores lógicos ya vistos. No obstante, ahora conviene aclarar algunos matices adicionales, en tanto que la existencia de cuantificadores trabajando sobre variables crea nuevas situaciones que deben ser analizadas con detenimiento.

1.4.4. Cálculo de predicados

Presentamos a continuación las propiedades más importantes que deben considerarse a la hora de trabajar con lógica de predicados.

- Sean A y B dos afirmaciones lógicas. Diremos que A **implica lógicamente** B si, dado cualquier universo U, se tiene que, para cualquier valor (o valores) de dicho universo, se cumple que B es verdadera siempre que A lo es.



Ejemplo

Sea A la afirmación “ $x > 4$ ” y la afirmación “ $x > 3$ ”. Entonces A implica lógicamente B, puesto que si un número es mayor que 4 automáticamente se puede concluir que debe ser mayor que 3. No obstante, B no implica lógicamente A, ya que un número que sea mayor que 3 no necesariamente ha de ser mayor que 4; por ejemplo, tomando $x = 4$, se tiene que $4 > 3$ pero, por el contrario, la afirmación “ $4 > 4$ ” es falsa.

- Sean A y B dos afirmaciones lógicas. Diremos que A **equivale lógicamente** a B si, dado cualquier universo U, se tiene que para cualquier valor (o valores) de dicho universo se cumple que A y B tienen siempre los mismos valores de verdad.

La definición anterior es equivalente a decir que A implica lógicamente B y, al mismo tiempo, B también implica lógicamente A.



Ejemplo

Sea A la afirmación “ $x > 5$ ” y B la afirmación “ $-x < -5$ ”. Entonces A es lógicamente equivalente a B, puesto que una relación se deduce a partir de la otra (en ambos sentidos), cambiando de signo la relación considerada.

Teorema

El siguiente resultado se conoce como las **leyes de Morgan generalizadas**, que consisten en una serie de equivalencias lógicas obtenidas a partir de los propios valores de verdad asociados los cuantificadores universal y existencial, expresándolos formalmente mediante afirmaciones conteniendo predicados.

1. $\neg[\forall x,p(x)] \leftrightarrow \exists x:\neg p(x)$.
2. $\neg[\exists x:p(x)] \leftrightarrow \forall x,\neg p(x)$.
3. $\forall x,p(x) \leftrightarrow \neg[\exists x,\neg p(x)]$.
4. $\exists x,p(x) \leftrightarrow \neg[\forall x,\neg p(x)]$.

Nota

Los cuantificadores de ambos tipos pueden combinarse de cualquier forma; no obstante, el orden es crucial, ya que una ordenación diferente puede cambiar por completo el significado del enunciado considerado y, con ello, su valor de verdad. Para ilustrar este hecho, consideremos el ejemplo siguiente, consistente en las dos afirmaciones que se presentan a continuación:

1. $\forall x \in N, \exists y \in N: y > x$.
2. $\exists y \in N: \forall x \in N, y > x$.

En el primer caso, el enunciado se lee como sigue: “para todo número natural, existe otro número natural mayor que el primero”. Es verdadero, puesto que, dado un $x \in N$, tomando por ejemplo $y = x + 1$, se tiene que $y \in N$ y, al mismo tiempo, $y > x$, puesto que tiene una unidad más que x .

No obstante, en el segundo caso, el enunciado se lee de la siguiente forma: “existe un número natural de forma que, para cualquier número natural, el primero siempre será mayor, sea cual sea el número natural elegido”. Esta vez, el enunciado es falso, ya que no hay ningún número natural que sea más grande que el resto. Además, en este caso el cuantificador universal admite un cambio de posición al final, quedando el enunciado de la siguiente forma:

el cual se lee como sigue: “existe un número natural mayor que cualquier otro número natural”, el cual es, efectivamente, equivalente al anterior (y, por tanto, también falso).

$$\exists y \in N: y > x, \forall x \in N,$$

En conclusión, el orden de los cuantificadores importa, puesto que puede modificar sustancialmente el enunciado original, si bien en algunos casos puede variarse.

Propiedad

Nos podemos encontrar con casos en los que algunas de las proposiciones que aparecen en un enunciado determinado pueden separarse del dominio de las variables cuantificadas por no depender de estas (y, por tanto, estar fuera del alcance de las mismas).

Si q es una proposición **futura del alcance de un cuantificador**, es decir, que no depende de la variable cuantificada, entonces cumple las propiedades siguientes, basadas en las equivalencias lógicas que se muestran a continuación:

1. $\forall x,[p(x) \vee q] \leftrightarrow [\forall x,p(x)] \vee q$.
2. $\exists x:[p(x) \vee q] \leftrightarrow [\exists x:p(x)] \vee q$.
3. $\forall x,[p(x) \wedge q] \leftrightarrow [\forall x,p(x)] \wedge q$.
4. $\exists x:[p(x) \wedge q] \leftrightarrow [\exists x:p(x)] \wedge q$.



Ejemplo

Las propiedades anteriores indican que, por ejemplo, es lo mismo afirmar “para todo número par, se cumple que este es además natural y que $2 + 2 = 4$ ”, que “todo número par es natural, y además $2 + 2 = 4$ ”, donde la primera afirmación se escribe como “ $\forall x \in \mathbb{N}, [p(x) \wedge q]$ ” y la segunda como “[$\forall x \in \mathbb{N}, p(x)$] \wedge q”, donde $p(x)$: “ser natural” y q : “ $2 + 2 = 4$ ”.

También cabe destacar las propiedades relativas a la asociatividad y distributividad asociadas a los cuantificadores.

Propiedad

Se cumplen las siguientes propiedades, consistentes en la equivalencia o implicación lógica que se muestran a continuación:

1. $[\forall x,p(x)] \vee [\forall x,q(x)] \rightarrow \forall x,[p(x) \vee q(x)]$.
2. $[\exists x: p(x)] \vee [\exists x: q(x)] \leftrightarrow \exists x:[p(x) \vee q(x)]$.
3. $[\forall x,p(x)] \wedge [\forall x,q(x)] \leftrightarrow \forall x,[p(x) \wedge q(x)]$.
4. $[\exists x: p(x)] \wedge [\exists x: q(x)] \rightarrow \exists x:[p(x) \wedge q(x)]$.



Ejemplo

Las propiedades anteriores permiten concluir que afirmaciones como “existe algún número tal que es múltiplo de 2 o bien es múltiplo de 3” y “existe algún número tal que es múltiplo de 2, o bien existe algún número tal que es múltiplo de 3” son equivalentes.

Nota

Obsérvese que, no obstante, en dos de los casos no se da la equivalencia, sino una relación de causalidad unidireccional (implicación). Esto se debe a que, efectivamente, en sentido inverso no necesariamente se cumplen, como veremos en los dos ejemplos que se muestran a continuación.

1. Consideremos la afirmación “todo número natural es par o bien impar”. Esta afirmación puede escribirse formalmente como $\forall x \in N, [p(x) \vee q(x)]$, donde $p(x)$: “ser par” y $q(x)$: “ser impar”. No obstante, la afirmación $[\forall x \in N, p(x)] \vee [\forall x \in N, q(x)]$ se lee como “todo número natural es par, o bien todo número natural es impar”, la cual es claramente diferente a la primera, ya que, de hecho, esta última es falsa (no todos los números naturales son pares ni todos los números naturales son pares), mientras que la segunda es verdadera.
2. Tomemos ahora el conjunto $A=\{1,2,3,4,5\}$ y la afirmación “existe algún número en A tal que es múltiplo de 2 y de 3 simultáneamente”, la cual puede escribirse formalmente como $\exists x \in A : [p(x) \wedge q(x)]$, siendo $p(x)$: “ser múltiplo de 2” y $q(x)$: “ser múltiplo de 3”. Sin embargo, la afirmación $[\exists x \in A : p(x)] \wedge [\exists x \in A : q(x)]$ se traduce como “existe algún elemento en A que es múltiplo de 2 y, a la vez, existe algún elemento en A tal que es múltiplo de 3” es diferente a la primera, ya que, de hecho, esta última es verdadera, mientras que la primera es falsa. En efecto, en el primer caso, no hay ningún número en A que sea simultáneamente múltiplo de 2 y de 3, mientras que sí que hay por separado algún múltiplo de 2 (por ejemplo, 2 y 4) y a la vez también hay algún múltiplo de 3 (el propio número 3).

Veamos a continuación una serie de ejemplos para poner en práctica estas nuevas nociones y propiedades establecidas.

Se propone la simplificación de las siguientes afirmaciones empleando todas las propiedades citadas anteriormente.

1. $\exists x : \neg[p(x) \wedge \neg q(x)]$.

Utilizando la ley de Morgan aplicada a proposiciones lógicas, obtenemos la equivalencia lógica siguiente:

$$\exists x : [\neg p(x) \vee \neg(\neg q(x))].$$

Ahora, empleando la propiedad relativa a la doble negación, obtenemos el resultado final:

$$\exists x : [\neg p(x) \vee q(x)].$$

La equivalencia entre el enunciado inicial y el enunciado final se hace patente cuando ambas afirmaciones se expresan verbalmente: en el primer caso, el enunciado se lee como “existe un x para el cual se da simultáneamente la ocurrencia de $p(x)$ y la no ocurrencia de $q(x)$ ”, mientras que en el segundo caso, la afirmación se lee como “existe un x cumpliendo que no se da $p(x)$ o bien se da $q(x)$ ”.

En este caso, también podríamos haber aplicado en un último paso la propiedad distributiva para obtener:

$$[\exists x: \neg p(x)] \vee [\exists x: q(x)].$$

2. $\neg[\forall x, [p(x) \vee q(x)]]$.

Empleando la correspondiente ley de Morgan generalizada, se obtiene la siguiente equivalencia lógica:

$$\exists x: \neg[p(x) \vee q(x)].$$

Ahora, utilizando la ley de Morgan aplicada a las proposiciones lógicas involucradas, se tiene:

$$\exists x: [\neg p(x) \wedge \neg q(x)].$$

Obsérvese que la equivalencia del enunciado original y del resultado final puede comprobarse verbalmente, ya que el primero se lee como “no es verdad que para todo x se cumpla que se tiene $p(x)$ o bien $q(x)$ ”, mientras que el último se lee como “existe algún x para el cual no se cumple $p(x)$ ni $q(x)$ ”.

En este caso, cabe hacer notar que no se podría haber aplicado la propiedad distributiva, en tanto que el cuantificador existencial es incompatible con la conjunción para tal efecto.

3. $\forall x, [\exists y: (p(x,y) \rightarrow q(x))]$.

Utilizando la equivalencia de la implicación lógica, se tiene que la expresión inicial es equivalente a la que se muestra a continuación:

$$\forall x, [\exists y: (\neg p(x,y) \vee q(x))].$$

Por otra parte, como \neg no depende de $,$, se tiene:

$$\forall x, [\neg p(x,y) \vee q(x)].$$

Obsérvese que, en este caso, tampoco podemos hacer uso de la propiedad distributiva, debido a la incompatibilidad del cuantificador universal con la disyunción para tal efecto.

4. $\exists x: [\forall y, \neg(p(x,y) \rightarrow q(x,y))]$.

En primer lugar, resolvemos el paréntesis interior, consistente en la implicación lógica:

$$\exists x: [\forall y, \neg(\neg p(x,y) \vee q(x,y))].$$

Ahora, aplicando la correspondiente ley de Morgan para la negación exterior, se tiene:

$$\exists x: [\forall y, (\neg(\neg p(x,y)) \wedge \neg q(x,y))].$$

Utilizando la doble negación en el primer argumento de la conjunción, obtenemos el siguiente resultado:

$$\exists x: [\forall y, (p(x,y) \wedge \neg q(x,y))].$$

En este caso, sí que estamos en condiciones de compatibilidad para aplicar la propiedad distributiva, por lo que, si se desea, el resultado se puede dejar como se muestra a continuación:

$$\exists x: [(\forall y, p(x,y)) \wedge (\forall y, \neg q(x,y))].$$

Por último, también es conveniente saber realizar el proceso inverso, es decir, traducir una afirmación de índole lógica expresada verbalmente a un enunciado formal que involucre predicados. Veamos a continuación algunos ejemplos que ilustran la forma de proceder en distintos casos.

Ejemplos:

- 1.** “Existe algún número negativo”.

Este enunciado se puede escribir de la siguiente forma:

$$\exists x, p(x),$$

siendo el predicado $p(x)$: “ser negativo” o, equivalentemente, “ $x < 0$ ”:

- 2.** “Todo múltiplo de 4 es par”.

En este caso, la afirmación considerada puede traducirse formalmente como sigue:

$$\forall x \in A, p(x),$$

siendo A el conjunto de todos los múltiplos de 4 y $p(x)$: “ser par”.

- 3.** “Todo número real es racional o bien irracional”.

En este caso, podemos escribir formalmente la afirmación anterior como se presenta a continuación:

$$\forall x \in R, [p(x) \vee q(x)],$$

donde $p(x)$: “ser racional” y $q(x)$: “ser irracional”.

Obsérvese que en este caso no puede aplicarse la propiedad distributiva, al tratarse de la combinación de un cuantificador universal y la disyunción entre dos proposiciones.

- 4.** “No es cierto que todo número primo sea impar, y además existen primos mayores que 100”.

La afirmación anterior se puede traducir formalmente de la siguiente manera:

$$\neg [\forall x \in P, p(x)] \wedge [\exists x \in P, q(x)],$$

donde $p(x)$: “ser impar” y $q(x)$: “ser mayor que 100” o, lo que es lo mismo, $x > 100$.

Aplicando la ley de Morgan generalizada correspondiente a la negación del primer argumento de la conjunción, la afirmación anterior puede traducirse como sigue:

$$[\exists x \in P: \neg p(x)] \wedge [\exists x \in P, q(x)].$$

En este caso, ambos argumentos de la conjunción no pueden unificarse mediante el cuantificador existencial, debido a su incompatibilidad con la conjunción para aplicar esta propiedad.

5. “Para todo número primo, sus divisores son, él mismo y ninguno más”.

Esta afirmación puede traducirse formalmente como sigue:

$$\forall x \in P, [p(x, 1) \wedge p(x, x) \wedge (\neg p(x, y), \forall y \notin A(x)],$$

siendo $p(x, y)$: “ x es divisible por y ” y $A(x)$ el conjunto formado por 1 y x , es decir, $A(x) = \{1, x\}$.

Dado que esta vez existe compatibilidad entre el cuantificador universal y la conjunción para aplicar la propiedad distributiva, la afirmación anterior puede escribirse como:

$$[\forall x \in P, p(x, 1)] \wedge [\forall x \in P, p(x, x)] \wedge [\neg p(x, y), \forall y \notin A(x)].$$

1.5. Lógica inductiva

Entendemos como lógica inductiva a aquella en la que, partiendo de premisas concretas, llegaremos a unas conclusiones más generales y amplias que no tendrán la garantía de ser ciertas.

Un **ejemplo** de ello sería:

- En todos los videojuegos que hemos jugado hasta el momento donde el enemigo era un orco, nuestro personaje siempre llevaba una espada.
- Por lo tanto, en todos los videojuegos con orcos como enemigos, el personaje que maneja el jugador portará una espada.

Dicho ejemplo nos sirve para tener una idea del mecanismo que sigue la lógica inductiva y como esta lógica “va de lo particular a lo general”, de una serie de casos o sucesos que siguen un patrón a una regla que pueda englobarlos y que como ya hemos dicho, no ha de ser tratada como una ley que pueda explicar todos los casos que estén por venir. Será la búsqueda de esta conclusión o norma general lo que trataremos más adelante.

1.5.1. Programación lógica inductiva: definición y ejemplos

Podemos definir la programación lógica inductiva como la intersección entre el aprendizaje inductivo con la programación lógica. Esto es, la zona de unión entre la hipótesis que vamos a crear a raíz de los casos que ya conocemos y el asentamiento de los nuevos casos que estén por llegar, junto con una representación más formal de los datos.



Ejemplo

Queremos pasar de una representación natural como:

El coche rojo ganará la carrera

A otra más formal y extrapolable a un modelo como:

$$C \equiv q \leftarrow A$$

Basándonos en una semántica que nos ayude a normalizarlos y valiéndonos de técnicas con los que poder llevarlos a un modelo computacional, podremos superar las limitaciones que nos surgen debido a la naturaleza de los datos.

- **Nota**

De ahora en adelante nos referiremos a la Programación Lógica Inductiva a través de sus siglas en inglés ILP.

Los sistemas ILP pueden clasificarse según varios criterios tales como:

- Según la cantidad de conceptos que se estén aprendiendo a la vez, llamados también predicados, pudiendo aprender uno o varios.
- Según la cantidad de ejemplos con los que empezar a funcionar:
 - Aprendizaje de una vez, donde todos los casos disponibles se tratarán al comienzo del proceso.
 - Aprendizaje secuencial, donde iremos agregando poco a poco aquellos casos disponibles.
- Según si es necesario su supervisión a lo largo del periodo de aprendizaje del sistema, reconocemos:
 - Interactivo: la persona a cargo revisa y corrige el proceso a lo largo de su periodo de aprendizaje.
 - No interactivo: cuando no tenemos a ninguna persona a cargo de la corrección del mismo.

Cada uno de estos criterios es independiente del anterior. Sin embargo, los ILP se suelen categorizar en dos tipos de sistemas:

- Sistemas de un único concepto, con aprendizaje de una sola vez y no interactivos, también llamados **ILP empíricos**.
- Sistemas de múltiples conceptos, con aprendizaje secuencial y aprendizaje interactivo, también llamados **ILP interactivos**.

Los sistemas ILP se encuentran en auge debido a las diversas aplicaciones que tienen en campos como el procesamiento del lenguaje natural, el control del tráfico o la biología molecular, por citar algunos.

Como ya hemos mencionado anteriormente, existen diversas limitaciones a las que atenernos cuando queremos representar nuestros datos para generar un sistema a partir de ellos. Como punto de partida sobre la representación formal de los mismos y de la dificultad que supone el manejo de lo que podríamos considerar el conocimiento desde el que nos basamos, podemos usar el formato Objeto-Atributo-Valor. Para el próximo ejemplo haremos uso tan solo de la última dupla, el Atributo-Valor.

Ejemplo:

- **Hipótesis** sobre la capacidad para aprobar un examen.

- **Conocimiento base:**

Si la dedicación al estudio de una asignatura es alta, existe una alta probabilidad de aprobar un examen.

Si la dedicación al estudio de una asignatura es media y dependiendo del hecho de que la dificultad del examen sea también acorde y de tipo test, entonces existe una alta probabilidad de aprobar dicho examen.

Si la dedicación al estudio de una asignatura es media y dependiendo del hecho de que la dificultad del examen sea más bien baja, entonces existe una alta probabilidad de aprobar dicho examen.

- Conjunción de **pares Atributo-Valor**:

(Dedicación=Media \wedge Dificultad=Media \wedge Tipo=Test)

V(Dedicación=Alta)

V(Dedicación=Media \wedge Dificultad=Baja)

Nota

Recordemos que el símbolo \wedge es el equivalente a una conjunción lógica, a la palabra ‘y’ o a una puerta lógica AND.

Asimismo tenemos que el símbolo V es el equivalente a una disyunción lógica, a la palabra ‘o’ o a una puerta lógica OR.

- Definición de las **reglas**:

- Si Dedicación=Media Y Dificultad=Media Y Tipo=Test \rightarrow Aprobar examen.
 - Si Dedicación=Alta \rightarrow Aprobar examen.
 - Si Dedicación=Media Y Dificultad=Baja \rightarrow Aprobar examen.

- Representación como **árbol de decisiones**.

Tomando el ejemplo anterior tendríamos:

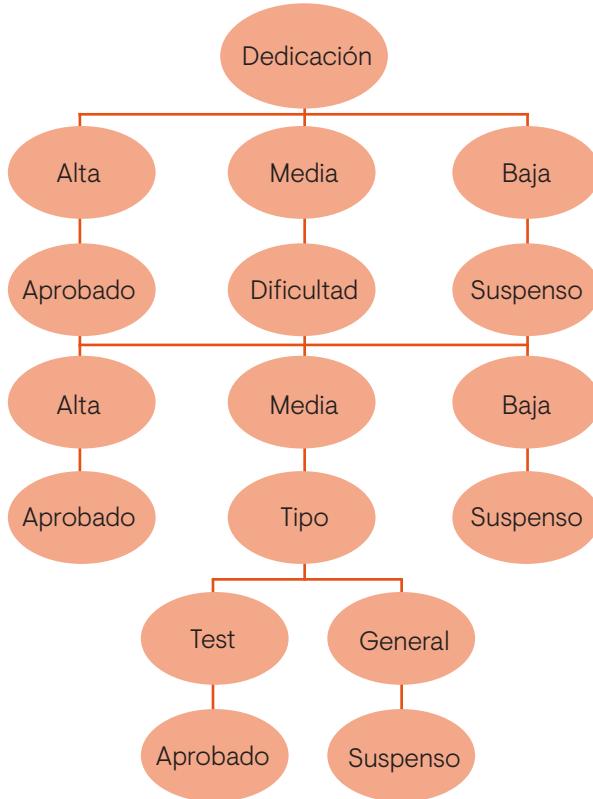


Figura 14. Representación gráfica de un árbol de decisiones. Fuente: elaboración propia.

1.5.2. Búsqueda de hipótesis

La búsqueda de hipótesis es la base de la programación lógica inductiva. Dentro del conjunto de todas las posibles hipótesis, tenemos que asegurarnos de dar cabida a las tuplas de datos que tenemos y que estas (las hipótesis) tienen un comportamiento acorde a ellas.

Las hipótesis las podemos categorizar dentro de una relación que va de menos a más generales cuantos más conjuntos de datos puedan albergar dentro de sí. Así mismo, en la búsqueda de dichas hipótesis podemos reconocer varias técnicas que nos puedan ayudar en nuestro propósito:

- Generalización relativa menos general

“Se basa en que si dos hipótesis son ciertas, la generalización más concreta de ambas también será mayormente cierta”.



Ejemplo

$C_1 = \text{Diabético (Ana)} \leftarrow \text{HijoDe (Ana,José)}, \text{PredisposGen (José)}$

$C_2 = \text{Diabético (Tomás)} \leftarrow \text{HijoDe (Tomás,Alvaro)}, \text{PredisposGen (Álvaro)}$

$H(C_1, C_2) = \text{Diabético (X)} \leftarrow \text{HijoDe (X,Y)}, \text{PredisposGen (Y)}$

- Resolución inversa

“Consiste en la inversión de las reglas de resolución, obteniendo así un método de generalización”.



Ejemplo

Tomando como base los siguientes hechos B

- $B_1 = \text{Diabético (Ana)}$
- $B_2 = \text{HijoDe (Ana,José)}$
- $B_3 = \text{PredisposGen (José)}$

Tomando y y usando la sustitución {Ana/X}

$C_1 = \text{Diabético (X)} \leftarrow \text{HijoDe (X,José)}$

Ahora haciendo uso de y y usando la sustitución { /Y}

$C_2 = \text{Diabético (X)} \leftarrow \text{HijoDe (X,Y)}, \text{PredisposGen (Y)}$

- Búsqueda en grafos refinados

“Se comienza con la hipótesis más general y se le aplican especializaciones hasta que solo cubren tuplas positivas”.

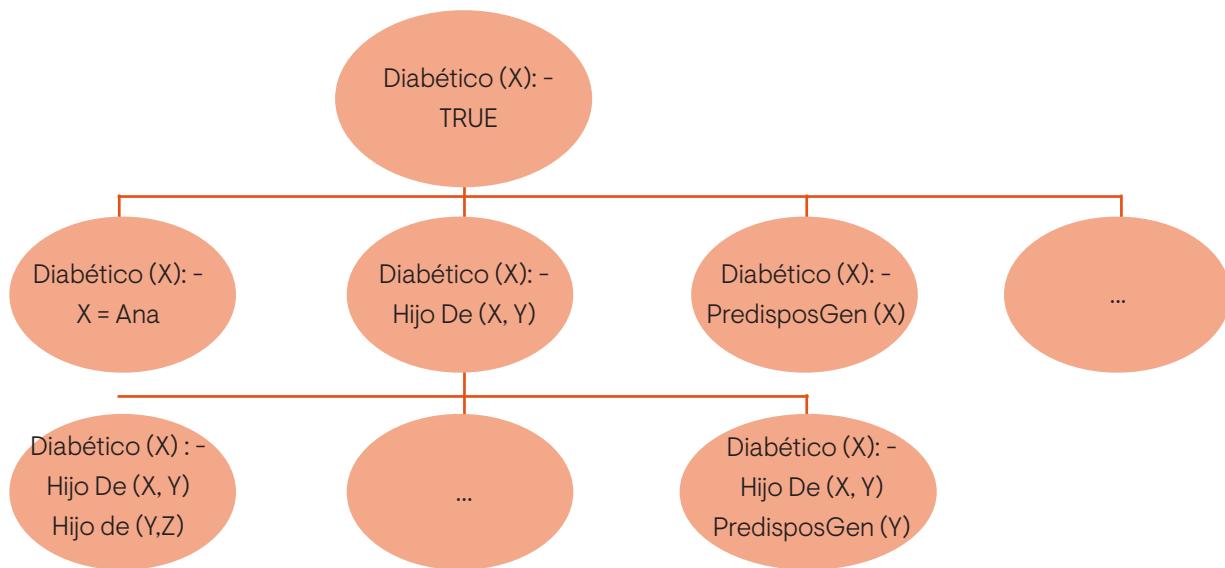


Figura 15. Representación una búsqueda de grafo refinado. Fuente: elaboración propia.

- Patrones de reglas

“Se denomina así debido a que las reglas son de un tipo especial en el que los predicados son variables”.

Ejemplo

Tomaríamos un patrón de reglas de la siguiente forma

$$P(X) \leftarrow Q(X, Y), R(Y)$$

Y tras aplicar los predicados que tenemos, se irían probando todas las combinaciones posibles con los conocimientos de base hasta dar con la regla adecuada.

$$\text{Diabético}(X) \leftarrow \text{Hijo De}(X, Y), \text{PredisposGen}(Y)$$

- Trasformación de problemas ILP a formato proposicional

“Se basa en la idea de convertir problemas de formato relacional a formato proposicional, utilizar sistemas de aprendizaje basados en lógica de objeto-atributo-valor y, finalmente, volver a convertir el conocimiento obtenido a formato relacional”.

Ejemplo

Clase	X	Y	HijoDe(X,X)	HijoDe(X,Y)	PredisposGen(Y)
+	Ana	José	FALSE	TRUE	TRUE
+	Tomás	Álvaro	FALSE	TRUE	TRUE
...					

Si el sistema de aprendizaje proposicional induce a la regla

[Clase +] si [Hijo De(X,Y) = TRUE] AND [PredisposGen(Y) = TRUE]

Entonces, podremos pasar nuevamente al formato relacional

Diabético (X) \leftarrow Hijo De (X,Y), PredisposGen (Y)

- Descubrimiento de cláusulas

O descubrimiento de hipótesis. Consiste en un enfoque diferente con respecto a lo que deberían ser los sistemas ILP. Dichos sistemas tienen por objetivo “la inducción de reglas que definan ciertos predicados. Estas reglas deberían ser lo suficientemente precisas como para poder reemplazar a los ejemplos de la base de datos y que estos fueran deducidos a partir de las mismas”. Sin embargo, en este caso “consiste en la búsqueda de restricciones dentro de la base de datos”.

Ejemplo

Ejemplo de reglas inducidas:

Padre De (Y,X) OR Madre De (Y,X) \leftarrow Hijo De (X,Y)
 Hijo De (X,Y) \leftarrow Padre De (Y,X)
 FALSE \leftarrow Hijo De(X,Y), PadreDe (X,Y)

En la última se deduce que una persona no puede ser hijo y padre de otra persona.

1.5.3. Inducción predictiva y descriptiva

Empezaremos por hablar sobre la inferencia empleando reglas, siendo esta una de las técnicas comunes en la minería de datos. Tenemos que, dependiendo del objetivo que busquemos con dicha inferencia, las podremos clasificar según sus técnicas en:

- **Inducción predictiva.** “Ámbito en el que podemos destacar el aprendizaje supervisado mediante reglas de clasificación. Para la construcción de las reglas de clasificación los algoritmos que generan los árboles de decisión siguen una estrategia de divide y vencerás, generando el árbol de arriba hacia abajo. Esto también se puede llevar a cabo en sentido contrario. Se puede tomar una clase y buscar la regla que cubre a todas las instancias de esa clase, excluyendo al mismo tiempo a aquellas instancias que no pertenecen a ella”.
- **Inducción descriptiva:** “Se destaca el descubrimiento de reglas de asociación siguiendo un modelo no supervisado de aprendizaje. El objetivo perseguido es encontrar reglas individuales que definan patrones interesantes en los datos. Las reglas de asociación se pueden obtener ejecutando un procedimiento de inducción de reglas de tipo divide y vencerás.

Sin embargo, un atributo de la parte derecha de la regla puede presentar cualquier valor y una única regla de asociación puede predecir en ocasiones el valor de más de un atributo. Para encontrar estas reglas habría que ejecutar el procedimiento de inducción de reglas una vez por cada posible combinación de atributos, con cada posible combinación de valores en la parte derecha de la regla.

Un nuevo mecanismo de inducción descriptiva es el denominado ‘descubrimiento de subgrupos’. En este dominio se lleva a cabo la búsqueda de subgrupos en el conjunto de datos que sean estadísticamente ‘más interesantes’, siendo tan grandes como sea posible y ofreciendo el mayor valor de atipicidad estadística con respecto a la propiedad en que estemos interesados. Se basa en obtener modelos descriptivos mediante el empleo de mecanismos predictivos”.

1.6. Teoría de complejidad computacional

Denominamos complejidad computacional a la magnitud de la complejidad de un algoritmo que da solución a un problema decidable o computable. Decimos que un problema es decidable o computable si las posibles respuestas del mismo son solo sí o no.

Para la resolución del problema mediante computación, vamos a necesitar dos tipos de recursos que van a medir la capacidad de realización del mismo, más allá de la optimización o no del algoritmo:

- Tiempo: que son la cantidad de pasos o veces que se tiene que ejecutar dicho algoritmo para dar con la solución.
- Espacio: Es la cantidad de memoria necesitada por el algoritmo para poder dar con la solución.



Ejemplo

A la vista de los dos recursos que hemos visto previamente, ¿cuál podríamos decir que es más valioso o crítico?

El tiempo va a ser el recurso que debamos tener más en cuenta, ya que el espacio a día de hoy no supone mucho problema en la computación moderna, además de ser un recurso reutilizable, cosa que no podemos decir del tiempo.

1.6.1. Definición de complejidad computacional: uso en el diseño de algoritmos

Cuando hablamos de complejidad computacional, no podemos hacerlo sin hacer referencia a los algoritmos utilizados para su resolución. Los algoritmos y el tiempo van de la mano ya que el modelo del algoritmo influirá mucho en este recurso.

Por ello, hablaremos de algoritmos de complejidad polinómica como aquellos que son tratables. Por un lado, tenemos que el problema tendrá que tener una solución de forma polinómica y que dicho problema deberá ser cerrado: es decir, existe un algoritmo que no solo es capaz de darle solución, sino que, además, es el más óptimo desde un punto de vista de los recursos disponibles.

Por otro lado, hemos comentado el concepto de tratable, el cual podríamos definir como la capacidad para poder llegar a la resolución en un sistema computacional. Este concepto es bastante subjetivo, ya que dependerá en buena medida de la naturaleza del problema y de la capacidad y los recursos con los que contemos.

Un ejemplo de problema tratable son los de búsqueda y los de ordenación. Así como existen problemas tratables, existen otros que podríamos calificar como intratables: aunque sean resolubles con una pequeña cantidad de datos de entrada, no son en absoluto escalables. Un ejemplo de estos últimos sería el conocido problema de las torres de Hanoi o el problema del viajero.

Ejemplo:

Un viajero quiere recorrer todas y cada una de las N ciudades recorriendo la menor distancia posible. Para poder hallar la solución a este problema tendremos que comprobar todas y cada una de las posibles combinaciones, siendo la solución a este problema = $N!$ De forma que para $N = 10$ tenemos 3 628 800 posibles recorridos, siendo esta cifra cada vez más inmanejable conforme aumentamos el número de ciudades disponibles.

Así como hemos mencionado la complejidad de un problema debido a la cantidad de sus datos de entrada, podemos definir criterios de escalabilidad en función de los recursos que consuman:

- Cantidad de tiempo en función del número de entradas n : $T(n)$
- Cantidad de espacio en función del número de entradas n : $S(n)$

Ejemplo de distintos tipos de funciones de algoritmos por entrada y su coste:

n función	10	50	100	300	1000
$5n$	50	250	500	1500	5000
$n \log_2 n$	33	282	665	2469	9966
n^2	100	2500	10000	90000	un millón (7 díg.)
n^3	1000	1250000	un millón (7 díg.)	27 millones (8 díg.)	mil millones (10 díg.)
2^n	1024	16 díg.	31 díg.	91 díg.	302 díg.
$n!$	3600000 (7 díg.)	65 díg.	161 díg.	623 díg.	inimaginable
n^n	11 díg.	85 díg.	201 díg.	744 díg.	inimaginable

Figura 16. Representación del coste por número de entradas dependiendo del tipo de función. Fuente: Pablo R. Fillottrani.

1.6.2. Modelos de computación

Podemos clasificar los problemas a solucionar en base a dos criterios diferenciados: la **Teoría de la Computabilidad** y la **Teoría de la Complejidad Computacional**.

Antes de seguir avanzando, nos vamos a detener a comentar el modelo de la máquina de Turing: “Este modelo fue definido por Alan Turing y aunque se puede pensar en la máquina de Turing como un ordenador moderno simplificado, especialmente en el caso determinista, surgió (y en muchas ocasiones será preferible considerarla en este sentido) como un modo de formalizar los algoritmos, para poder razonar matemáticamente sobre ellos. De hecho, uno de los motivos de su sencillez es que modeliza la idea de una persona haciendo cálculos provista de lápiz y un borrador. De esta forma, cada máquina resuelve un problema, representa un algoritmo”.

Nota

De ahora en adelante nos referiremos al modelo de máquina de Turing como MT.

En base a la Teoría de la Computabilidad, clasificaremos los problemas de forma que sus respuestas serán SI/NO, VERDADERO/FALSO o ‘0’/‘1’, para nos ayude a tomar decisiones:

- Decidible o resoluble algorítmicamente:
 - Si existe un procedimiento mecánico (MT) que lo resuelva.
 - Además, la MT debe detenerse para cualquier entrada.
- Parcialmente decidible (reconocible):
 - Si existe un procedimiento mecánico (MT) que lo resuelva.
 - Además, la MT debe detenerse para aquellas entradas que son una solución correcta al problema.
- No decidible:
 - Si NO es decidible.

En base a la Teoría de la Complejidad Computacional, clasificaremos los problemas a resolver en las siguientes clases: Clase L, NL, P, P-Completo, NP, NP-Completo, NP-Duro...

Asociadas a dichas clases, está la máquina de Turing dentro de la cual debemos diferenciar los siguientes tipos:

- **MT determinista:** para este tipo de maquina solamente existe una posible transición, esto es, de un estado a otro.
- **MT no determinista:** para este tipo de maquina existe más de una posible transición, esto es, de un estado puede cambiar a varios diferentes.

1.6.3. Clases de complejidad

En este apartado nos vamos a centrar en tres de las clases anteriormente mencionadas: P, NP y NP-Completo. Sin embargo, antes vamos a examinar un nuevo concepto: se denomina tiempo polinómico a aquel tiempo de ejecución de un algoritmo. Por lo tanto, cuando decimos que un algoritmo se puede ejecutar en tiempo polinómico, estamos diciendo que es menor que un cierto valor que consideraremos asumible o de calidad.

- **Clase P**

Se trata de problemas con MT determinista y resolubles en tiempo polinómico. Dichos problemas son tratables, esto es, el tiempo de ejecución para hallar la solución se encuentra dentro de lo que se considera razonable. Un tipo de problemas de esta clase son los de ordenación y los de búsqueda.

- **Clase NP**

Se trata de problemas con MT no determinista y resolubles en tiempo polinómico. Los MTD son considerados un caso particular de MTND con lo que podemos considerar la clase P como un subconjunto de la clase NP.

- **Clase NP-Completo**

Consideramos un problema de clase NP-Completo si es NP y, además, todos los demás problemas de NP pueden derivar en este. De esta forma, la solución hallada para uno, puede ser utilizada para los demás.

1.7. Teoría de grafos

La teoría de grafos es una herramienta que diseñó el célebre matemático suizo Leonhard Euler para resolver un problema clásico de una forma elegante y efectiva. En este apartado, introduciremos los aspectos más relevantes de esta teoría y veremos algunas de entre sus muchas aplicaciones para abordar problemas determinados, cuya resolución sería complicada sin disponer de dicha teoría. Un libro interesante que puede utilizarse como complemento es el de Calcedo, Wagner de García, Méndez (*Introducción a la teoría de grafos*).

Grafo

Un grafo es una terna, (V, E, γ) , donde V y E son conjuntos, así como una aplicación:

$$\gamma : E \rightarrow \mathcal{E},$$

donde $\mathcal{E} = \{\{v, v'\} \mid v, v' \in V\}$.

El conjunto recibe el nombre de conjunto de **vértices**, el conjunto recibe el nombre de conjunto de **lados o aristas** y la aplicación γ_G , **aplicación de incidencia**. La finalidad de la aplicación es interpretar un lado o arista como un nexo entre dos vértices (razón por la cual se le asigna un conjunto de vértices).

Los grafos pueden **representarse gráficamente** a partir de su definición, como veremos en los siguientes ejemplos.

Ejemplos:

1. Consideraremos el grafo $G = (V, E, \gamma)$ con conjunto de vértices $V = \{v_1, v_2, v_3, v_4\}$ y conjunto de aristas dado por $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$, siendo $\gamma : E \rightarrow \mathcal{E}$ la aplicación dada por:

$$\begin{aligned}
 \gamma(e_1) &= \{v_1, v_2\} \\
 \gamma(e_2) &= \{v_1, v_2\} \\
 \gamma(e_3) &= \{v_2, v_3\} \\
 \gamma(e_4) &= \{v_2, v_3\} \\
 \gamma(e_5) &= \{v_1, v_4\} \\
 \gamma(e_6) &= \{v_2, v_4\} \\
 \gamma(e_7) &= \{v_3, v_4\}
 \end{aligned}$$

Entonces el grafo G puede representarse gráficamente como se muestra a continuación.

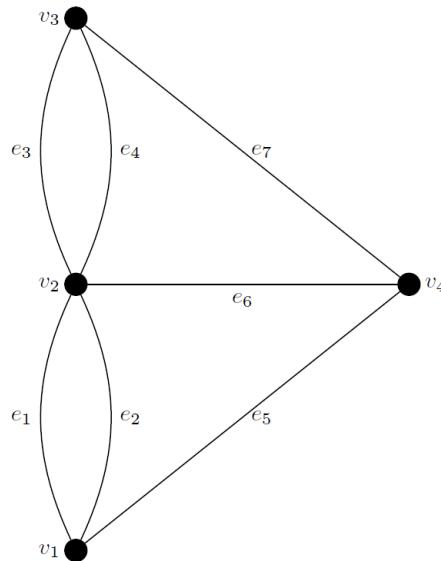


Figura 17. Representación gráfica del grafo G. Fuente: elaboración propia.

2. Sea el grafo $G=(V,E,\gamma)$, dado por el conjunto de vértices $V=\{v_1, v_2, v_3\}$ y el conjunto de aristas $E=\{e_1, e_2, e_3, e_4\}$, con $\gamma:E \rightarrow \mathcal{E}$ la aplicación definida como:

$$\begin{aligned}
 \gamma(e_1) &= \{v_1, v_1\} = \{v_1\} \\
 \gamma(e_2) &= \{v_1, v_2\} \\
 \gamma(e_3) &= \{v_1, v_3\} \\
 \gamma(e_4) &= \{v_2, v_3\}
 \end{aligned}$$

Entonces, podemos representar gráficamente el grafo G como sigue:

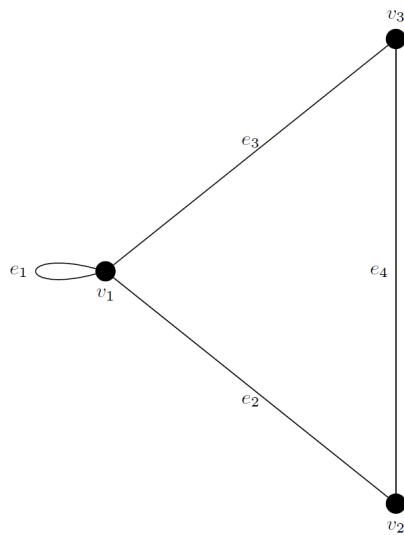


Figura 18. Representación gráfica del grafo . Fuente: elaboración propia.

Subgrafo

Sea $G=(V,E,\gamma)$ un grafo. Diremos que $H=(V',E',\gamma')$ es un subgrafo G si $V' \subseteq V$, $E' \subseteq E$ y además $\gamma':E' \rightarrow \mathcal{E}'$ es tal que $\gamma'(e)=\gamma(e), \forall e \in E'$, donde

$$\mathcal{E}' = \{\{v, v'\} | v, v' \in E'\}$$

Ejemplos:

1. Sea G el grafo considerado en el primer ejemplo presentado anteriormente. Entonces $H=(V',E',\gamma')$, dado por $V'=\{v_1, v_2, v_3\}$ y $E'=\{e_1, e_2, e_3, e_4\}$, con $\gamma':E' \rightarrow \mathcal{E}'$ la aplicación dada por:

$$\begin{aligned}\gamma(e_1) &= \{v_1, v_2\} \\ \gamma(e_2) &= \{v_1, v_2\} \\ \gamma(e_3) &= \{v_3, v_3\} \\ \gamma(e_4) &= \{v_2, v_3\}\end{aligned}$$

es un subgrafo de G.

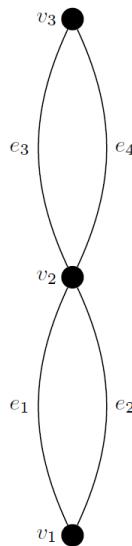


Figura 19. Representación gráfica del subgrafo H. Fuente: elaboración propia.

Tal y como puede comprobarse en su representación gráfica, el subgrafo H consiste en eliminar el vértice del grafo original y, con él, todas las aristas adyacentes a este, es decir, e_5, e_6 y e_7 .

2. Sea G el grafo considerado en el segundo ejemplo presentado anteriormente. Entonces $H = (V', E', \gamma')$, dado por $V' = \{v_1, v_2, v_3\}$ y $E' = \{e_2, e_3, e_4\}$, siendo $\gamma': E \rightarrow E'$ la aplicación dada por:

$$\gamma(e_2) = \{v_1, v_2\}$$

$$\gamma(e_3) = \{v_1, v_3\}$$

$$\gamma(e_4) = \{v_2, v_3\}$$

es un subgrafo de G.

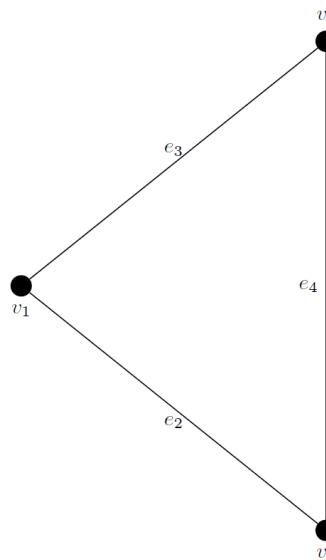


Figura 20. Representación gráfica del subgrafo . Fuente: elaboración propia.

En este caso, puede observarse que el subgrafo consiste en eliminar la arista e_1 del grafo original, que va del vértice v_1 hacia sí mismo.

Grafo simple

Un grafo $G=(V,E,\gamma)$ simple es un grafo tal que cumple dos restricciones adicionales:

1. No hay aristas diferentes que unan el mismo par de vértices, es decir, si $e \neq e'$, entonces $\gamma(e) \neq \gamma(e')$.
2. No contiene bucles o *loops*, es decir, no tiene aristas que unen un vértice consigo mismo o, lo que es lo mismo, $\gamma(e)=\{v_e, v_{e'}\}$ cumple que $v_e \neq v_{e'}, \forall e \in E$.

Ejemplos:

1. El grafo del primer ejemplo anterior no es simple, ya que tiene varias aristas que unen el mismo par de vértices, por ejemplo, el par de aristas $\{e_1, e_2\}$ y el par de aristas $\{e_3, e_4\}$
2. El grafo del segundo ejemplo anterior tampoco es simple, en tanto que la primera arista consiste en bucle, pues $\gamma(e_1)=\{v_1, v_1\}=\{v_1\}$.
3. Tomemos el grafo $G=(V,E,\gamma)$, dado por el conjunto de vértices $V=\{v_1, v_2, v_3, v_4\}$, el conjunto de aristas $E=\{e_1, e_2, e_3, e_4\}$ y $\gamma: E \rightarrow \mathcal{E}$ la aplicación dada por:

$$\gamma(e_1) = \{v_1, v_2\}$$

$$\gamma(e_2) = \{v_1, v_3\}$$

$$\gamma(e_3) = \{v_2, v_3\}$$

$$\gamma(e_4) = \{v_3, v_4\}$$

Entonces, este grafo es simple (ya que aristas diferentes unen pares de vértices diferentes y no hay bucles) y su representación gráfica es la que se muestra seguidamente.

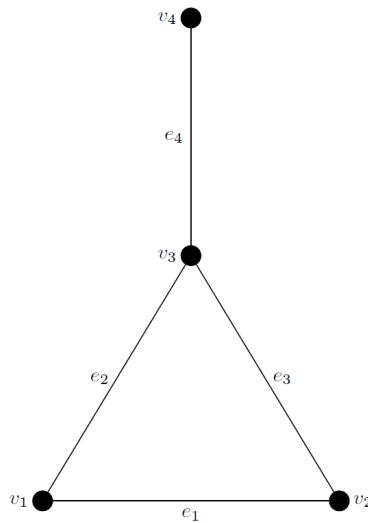


Figura 21. Representación gráfica del grafo simple . Fuente: elaboración propia.

En ocasiones, interesa utilizar grafos para representar caminos que solo pueden recorrerse de un vértice a otro, pero no en sentido opuesto. Para ello, debemos dotar a la definición original de grafo de una nueva estructura que permita hacer tal distinción. Veamos a continuación la definición más adecuada para ello.

Grafo dirigido

Un grafo dirigido G es una terna (V, E, γ) , donde V y E son conjuntos (de vértices y de aristas o lados, respectivamente) e γ es una aplicación de la forma:

$$\gamma : E \rightarrow V \times V$$

Los grafos dirigidos habitualmente se representan gráficamente dibujando las aristas con una flecha en la dirección válida para recorrerla, siendo la primera componente del par ordenado el punto de partida y la segunda componente, el punto de llegada.

Nota

Obsérvese que la única diferencia entre la noción de grafo y la de grafo dirigido es que en el primer caso se utiliza un conjunto para representar los pares y, en el segundo caso, se utiliza un par ordenado, por lo que en ese caso el orden sí que importa. Ahí estriba, por tanto, la diferencia clave entre ambos conceptos.

Ejemplo:

Si variámos el último ejemplo anterior, es decir, el grafo G con el conjunto de vértices $V=\{v_1, v_2, v_3, v_4\}$, el conjunto de aristas $E=\{e_1, e_2, e_3, e_4\}$ y reemplazando esta vez la definición de γ por $\gamma: E \rightarrow V \times V$ dada por:

$$\begin{aligned}\gamma(e_1) &= \{v_1, v_2\} \\ \gamma(e_2) &= \{v_1, v_3\} \\ \gamma(e_3) &= \{v_2, v_3\} \\ \gamma(e_4) &= \{v_3, v_4\}\end{aligned}$$

Si representamos gráficamente este grafo, obtenemos la figura que se muestra a continuación.

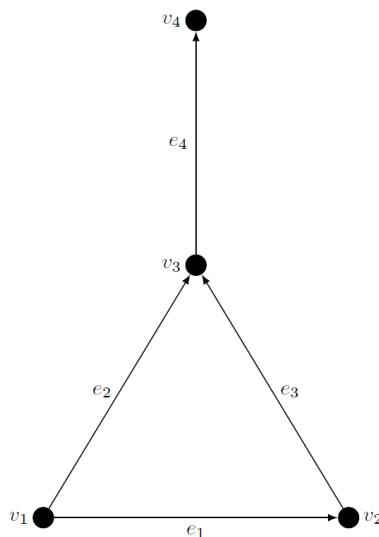


Figura 22. Representación gráfica del grafo dirigido G . Fuente: elaboración propia.

Nota

En adelante, por comodidad, omitiremos la existencia de la aplicación e identificaremos directamente cada arista con el conjunto o par ordenado (según sea un grafo dirigido o no dirigido, respectivamente) de vértices al cual está asociada por dicha aplicación.

Matriz de adyacencia

Sea G un grafo simple (pero admitiendo la posible existencia de bucles) con n vértices, $V=\{v_1, v_2, \dots, v_n\}$. La matriz de adyacencia asociada a G es una matriz cuadrada y simétrica $n \times n$, denotada por A , tal que:

- $A(i,j) = 1$, si los vértices v_i y v_j están unidos por una arista.
- $A(i,j) = 0$, en caso contrario.

Nota

Las matrices de adyacencia determinan únicamente un único grafo simple, caracterizándolo por completo. Por tanto, a partir de un grafo simple puede obtenerse una matriz de adyacencia y, recíprocamente, dada una matriz de adyacencia puede determinarse únicamente el grafo al cual representa dicha matriz.

Por otra parte, las matrices de adyacencia también pueden definirse para grafos dirigidos, con la salvedad de que en este caso la matriz resultante no será simétrica.

Ejemplos:

1. La matriz de adyacencia asociada al grafo del ejemplo anterior es:

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

2. Sea G el grafo dado por 5 vértices con aristas:

$$E = \{\{1,2\}, \{1,3\}, \{2,4\}, \{3,4\}, \{3,5\}, \{4,5\}\}.$$

Entonces, G puede representarse gráficamente como sigue:

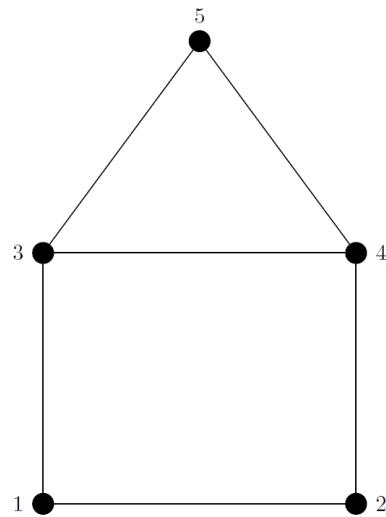


Figura 23. Representación gráfica del grafo G . Fuente: elaboración propia.

3. Tomemos ahora la matriz de adyacencia dada por:

$$A = \begin{vmatrix} 111 \\ 101 \\ 110 \end{vmatrix}$$

El grafo asociado tiene 3 vértices con aristas $E=\{\{1,1\},\{1,2\},\{1,3\},\{2,3\}\}$. Gráficamente, el grafo tiene el aspecto que se muestra a continuación:

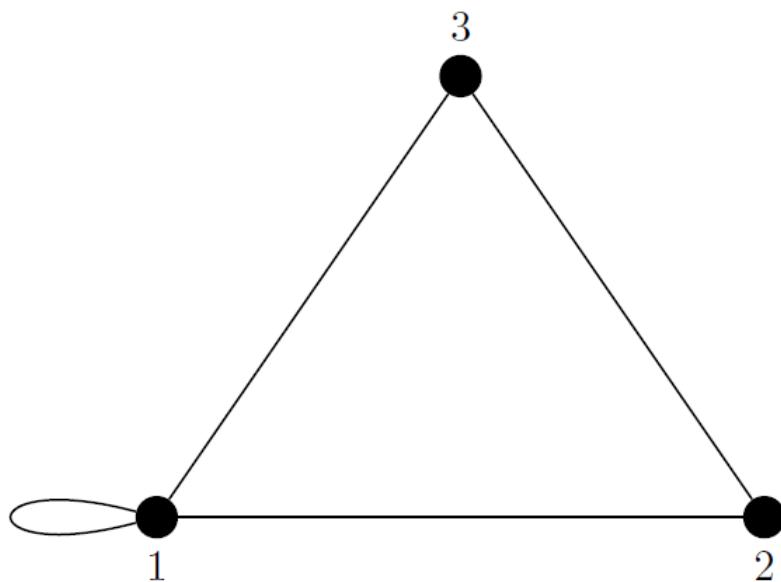


Figura 24. Representación gráfica del grafo definido a partir de la matriz de adyacencia A . Fuente: elaboración propia.

4. Sea G el grafo dirigido de 4 vértices dado por las aristas $(1,2)$, $(1,3)$, $(2,1)$, $(2,4)$, $(3,4)$. Podemos representarlo gráficamente como sigue:

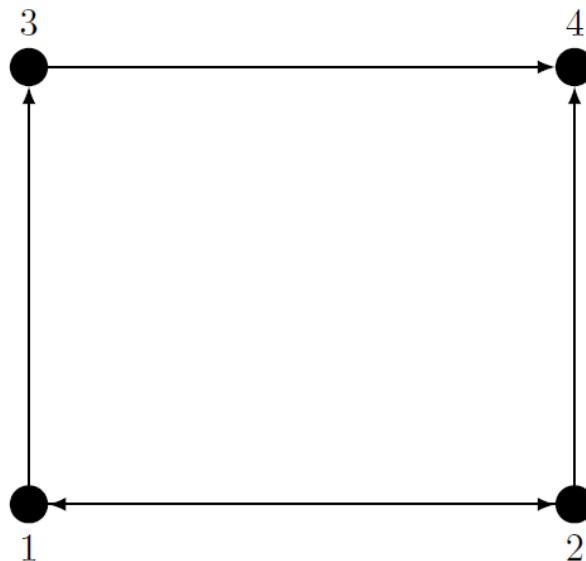


Figura 25. Representación gráfica del grafo dirigido G . Fuente: elaboración propia.

Su matriz de adyacencia es, por tanto:

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

1. Caminos y conexión

Camino

Sea G un grafo. Un camino C entre dos vértices v_1 y v_n es una sucesión de aristas de la forma $\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}$ que une los vértices v_1 y v_n .

Además, dado un camino entre v_1 y v_n , se tiene:

- El número de aristas que recorre se denomina **longitud** del camino.
- El camino se llama **cerrado** si $v_1 = v_n$ y abierto, en caso contrario.
- Se dice que es **simple** si no pasa dos veces por el mismo vértice.
- Es un **ciclo** si es simple y cerrado.

Ejemplo:

Consideremos el grafo de 4 vértices con aristas $\{1,2\}$, $\{1,3\}$, $\{1,4\}$, $\{2,4\}$, $\{3,4\}$. Gráficamente, puede representarse como sigue.

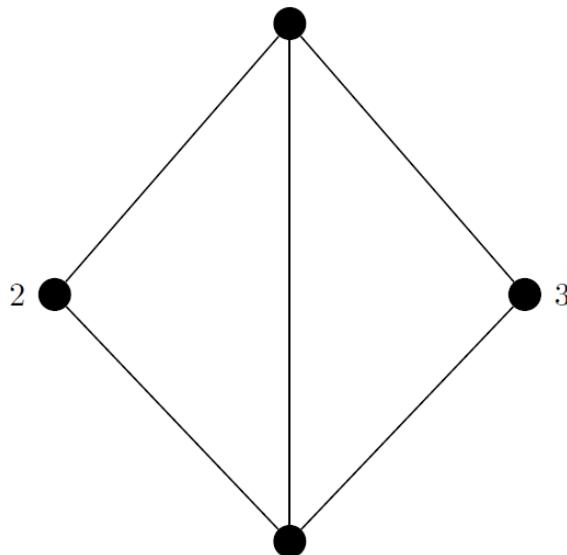


Figura 26. Representación gráfica. Fuente: elaboración propia.

Vamos a considerar distintos caminos para este grafo.

- $\{2,4\}, \{4,3\}$ es un camino entre 2 y 3, que cumple:
 - Tiene longitud 2.
 - Es abierto.
 - Es simple.
 - No es un ciclo.
- $\{2,4\}, \{4,1\}, \{1,3\}$ es un camino entre 2 y 3, cumpliendo:
 - Tiene longitud 3.
 - Es abierto.
 - Es simple.
 - No es un ciclo.
- $\{1,2\}, \{2,4\}, \{4,1\}$ es un camino entre 1 y 1, cumpliendo:
 - Tiene longitud 3.
 - Es cerrado.
 - Es simple.
 - Es un ciclo.

- $\{2,4\}, \{4,1\}, \{4,3\}, \{3,4\}, \{4,2\}$ es un camino entre 2 y 2, que cumple:

- Tiene longitud 5.
- Es cerrado.
- No es simple.
- No es un ciclo.

Conexión

Un grafo G es conexo si para cada par de vértices diferentes existe algún camino que los une.

Ejemplos:

1. Sea G el grafo de 5 vértices dado por las aristas $\{1,2\}, \{1,3\}, \{2,3\}, \{3,4\}, \{3,5\}, \{4,5\}$. Entonces G es conexo, ya que todos los vértices tienen un camino, es decir, una sucesión de aristas adjuntas, que los conectan. Puede comprobarse representándolo gráficamente.

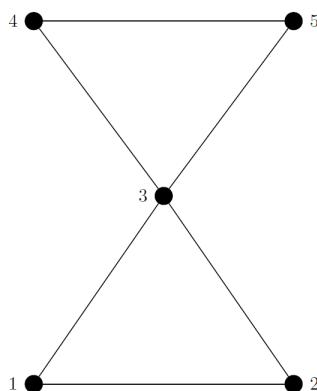


Figura 27. Representación gráfica del grafo conexo G . Fuente: elaboración propia.

2. Sea G el grafo de 5 cuyas aristas son $\{1,2\}, \{1,3\}, \{2,3\}, \{4,5\}$. Entonces G no es conexo, ya que, por ejemplo, para los vértices 3 y 4 no hay ningún camino que los una. Este hecho puede comprobarse gráficamente con mayor facilidad.

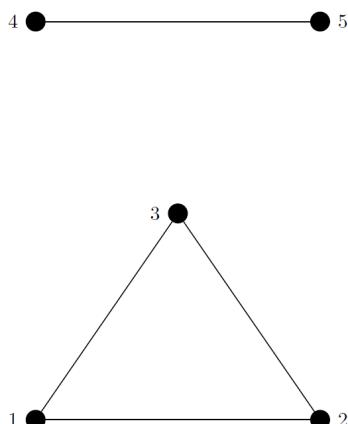


Figura 28. Representación gráfica del grafo no conexo G . Fuente: Elaboración propia.

Camino euleriano

Sea G un grafo. Un camino euleriano de G es un camino tal que recorre todos sus vértices pasando una única vez por todas y cada una de sus aristas. Si, además, dicho camino es cerrado, entonces recibe el nombre de **ciclo euleriano**.

Ejemplos:

1. Sea el grafo de 5 vértices cuyas aristas son $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{3,4\}$, $\{3,5\}$, el cual puede representarse gráficamente como sigue:

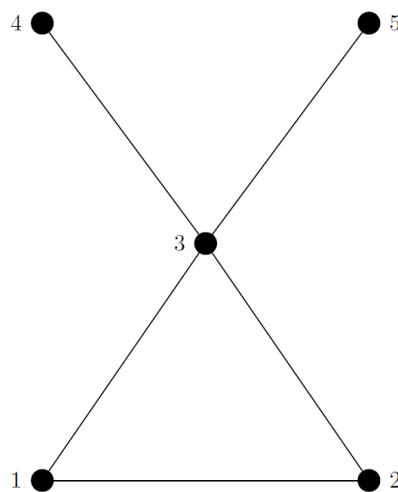


Figura 29. Representación gráfica del grafo G . Fuente: elaboración propia.

Entonces G tiene un camino euleriano; por ejemplo, tomemos la secuencia de aristas $\{4,3\}$, $\{3,2\}$, $\{2,1\}$, $\{1,3\}$, $\{3,5\}$, un camino que cumple con la condición de camino euleriano.

2. Tomamos ahora el grafo de 5 vértices dado por las aristas $\{1,2\}$, $\{1,3\}$, $\{2,3\}$, $\{3,4\}$, $\{3,5\}$, $\{4,5\}$, el cual podemos representar gráficamente de la siguiente forma:

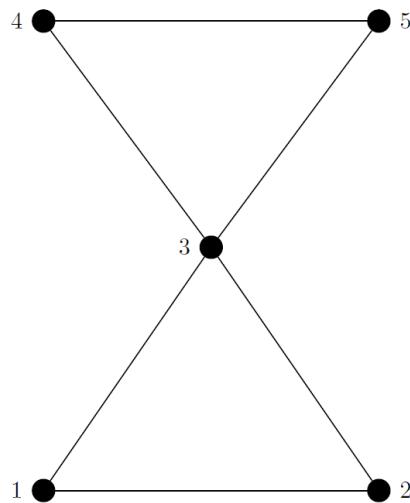


Figura 30. Representación gráfica del grafo G . Fuente: elaboración propia.

Entonces G tiene un ciclo euleriano; en efecto, consideramos, por ejemplo, la secuencia de aristas $\{3,4\}, \{4,5\}, \{5,3\}, \{3,1\}, \{1,2\}, \{2,3\}$, la cual es un camino que satisface la condición de ciclo euleriano.

Interesa saber si existe algún procedimiento general que permita determinar cuándo un grafo tiene un camino euleriano, cuándo tiene un ciclo euleriano o cuándo no puede tener alguno de los dos elementos anteriores. Para ello, primero debemos introducir el concepto que se define a continuación.

Grado

Sea G un grafo y $v \in V$ un vértice. El grado de v , denotado por $g(v)$, es el número de aristas incidentes a dicho vértice. Dicho de otra forma, representa el número de conexiones que tiene con otros vértices.

Ejemplos:

1. El grado asociado a cada uno de los vértices del grafo del primer ejemplo anterior es el siguiente:

- $g(1) = 2$.
- $g(2) = 2$.
- $g(3) = 4$.
- $g(4) = 1$.
- $g(5) = 1$.

2. El grado de cada uno de los vértices del grafo del segundo ejemplo anterior es el siguiente:

- $g(1) = 2$.
- $g(2) = 2$.
- $g(3) = 4$.
- $g(4) = 2$.
- $g(5) = 2$.

Una vez introducido el concepto de grado de un vértice, ya estamos en condiciones de presentar el resultado que caracteriza la existencia de caminos y ciclos eulerianos.

Teorema

Sea G un grafo. Entonces se cumplen:

- G tiene algún camino euleriano si y solo si tiene exactamente dos vértices de grado impar.
- G tiene algún ciclo euleriano si y solo si todos sus vértices tienen grado par.

Además, si tiene un camino euleriano, entonces necesariamente sus dos extremos deben ser sendos vértices de grado impar.

Ejemplos:

1. El grafo del primer ejemplo anterior tiene un camino euleriano, ya que tiene exactamente dos vértices de grado impar (además de porque ya hemos encontrado uno explícitamente), pero no tiene ningún ciclo euleriano, ya que no todos sus vértices son de grado impar.
2. El grafo del segundo ejemplo tiene un ciclo euleriano, puesto que el grado de todos sus vértices es par (y, además, porque ya hemos encontrado previamente tal ciclo), pero no tiene ningún camino euleriano, en tanto que no tiene exactamente dos vértices de grado impar, sino que de hecho no tiene ninguno con tal característica.
3. Sea G el grafo de cuatro vértices dado por las aristas $\{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}$, el cual puede representarse gráficamente de la siguiente manera:

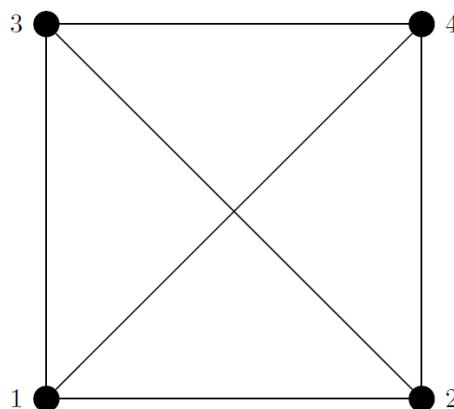


Figura 31. Representación gráfica de G . Fuente: elaboración propia.

Puede comprobarse que, en este caso, se tiene:

- $g(1) = 3$.
- $g(2) = 3$.
- $g(3) = 4$.
- $g(4) = 3$.
- $g(5) = 3$.

Por tanto, no se cumplen ni las condiciones para tener un camino euleriano ni las condiciones para poder tener un ciclo euleriano, ya que el número de vértices de grado impar es cuatro. Por tanto, este grafo no tiene ni caminos ni ciclos eulerianos.

2. Árboles

Los árboles consisten en un tipo especial de grafo cuyo estudio permite resolver una gran variedad de problemas, entre ellos la determinación del camino más corto, el cual abordaremos a lo largo de esta sección.

2.1. Definiciones, ejemplos y árboles generadores

Un árbol es un grafo conexo y sin ciclos (caminos cerrados).

Ejemplos:

1. El siguiente grafo es un árbol, ya que es conexo y no presenta bucles.

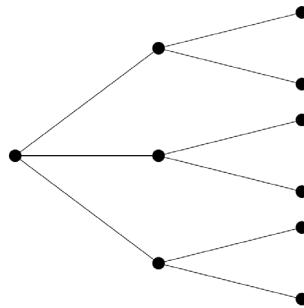


Figura 32. Ejemplo de un árbol. Fuente: elaboración propia.

2. El grafo que se muestra a continuación no es un árbol, ya que no es conexo.

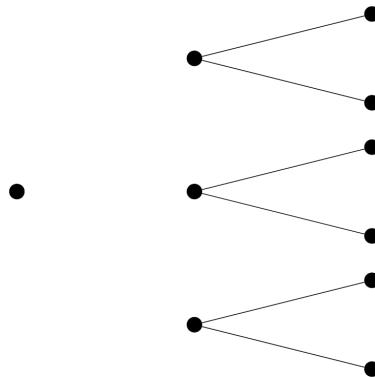


Figura 33. Este grafo no es un árbol por no ser conexo. Fuente: elaboración propia.

3. El grafo que se presenta seguidamente no es un árbol, en tanto que presenta ciclos.

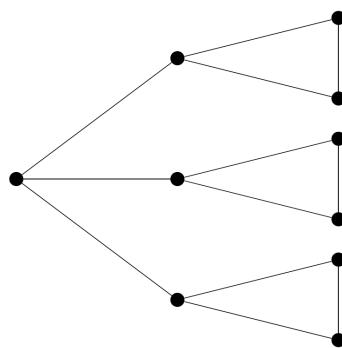


Figura 34. Este grafo no es un árbol, pues presenta tres ciclos. Fuente: elaboración propia.

Los árboles pueden caracterizarse de manera única a través del resultado que presentamos a continuación.

Teorema

Sea G un grafo. Entonces G es un árbol si y solo si $\forall v, v' \in V$, existe un único camino entre v y v' .

Árbol generador

Sea G un grafo. Diremos que un subgrafo de G , H , es un árbol generador de G si H es un árbol contenido en todos los vértices de G . Es decir, si H es un árbol y $V' = V$.

Ejemplo:

Se muestra a continuación un árbol generador asociado al primer grafo presentado en este tema.

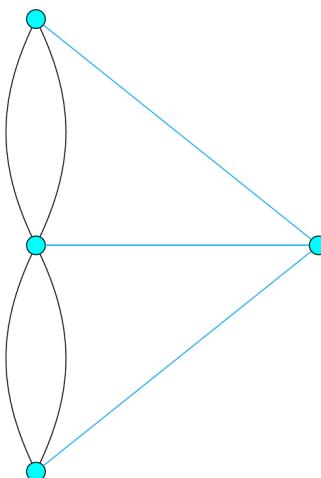


Figura 35. Ejemplo de grafo generador (con vértices y aristas resaltadas en color cian). Fuente: elaboración propia.

3. Algoritmo de Dijkstra: descripción y ejemplos de aplicación

En este apartado vamos a dedicarnos a abordar problemas relacionados con determinar el camino más corto entre dos puntos, dentro de una red con más puntos con conexiones complejas entre ellos. Este tipo de problemas pueden modelizarse mediante grafos, donde los nodos representan los lugares de partida o de destino y las aristas sus correspondientes conexiones.

No obstante, en esta ocasión también debemos considerar un ingrediente más, que es la **longitud** de dichas conexiones o caminos entre cada uno de los vértices o nodos conectados entre sí. Por ello, consideraremos lo que se conoce como **grafos ponderados**, que consisten en grafos en el sentido habitual, donde, además, se considera para cada arista un **peso**, asociado a su longitud. Formalmente, los pesos asociados a las aristas de un grafo $G = (V, E, y)$ pueden considerarse como una aplicación $p : E \rightarrow [0, +\infty]$, es decir, donde a cada arista se le asigna un peso consistente en un valor real positivo.

Teniendo en cuenta estas consideraciones, presentamos a continuación el algoritmo de Dijkstra, consistente en hallar el camino más corto entre dos vértices de un grafo conexo y ponderado. En términos de ponderación, esto quiere decir encontrar el camino entre los dos vértices cuya suma de los pesos de las aristas sea la mínima posible.

Algoritmo

Supongamos que tenemos un grafo con n vértices, listados como $\{v_1, v_2, \dots, v_n\}$ y queremos hallar el camino más corto entre v_1 y v_n . El algoritmo consiste en la generación de un árbol que, en última instancia, contendrá el camino más corto entre v y v_n .

Para ello, se van generando los pesos de los caminos mínimos desde el vértice inicial, v , hasta los demás vértices, cuyos datos asociados se van almacenando en una tabla con una estructura similar a ésta:

Iteración	$p(v_1)$	$p(v_2)$...	$p(v_n)$	V	E
:		:			:	:

En la columna V se van añadiendo los vértices que van definiendo el árbol a partir de la aplicación del algoritmo, mientras que en la columna E se añaden las aristas que se van eligiendo, las cuales definen en última instancia el árbol que determinará el camino más corto (el único camino que tendrá entre v_1 y v_n).

Dicha tabla se va llenando a partir del seguimiento de los pasos siguientes:

- Definimos $p(v_1) = 0$, lo que representa que nos encontramos en el punto de partida (la distancia de v_1 con respecto a sí mismo es 0).
- En la primera iteración, para cada vértice v_j , $2 \leq j \leq n$, si v_j es adyacente a v_1 , entonces, es decir, $p(v_j)$ se define como el peso asociado a la arista que los une. En otro caso, $p(v_j) = \infty$.

De ese modo, el valor mínimo del conjunto $\{p(v_2), \dots, p(v_n)\}$ proporciona el vértice más próximo a v_1 . Supongamos que llamamos x a dicho vértice. Entonces, en la columna V se añade el vértice x , a la columna E la arista $\{v_1, x\}$ y el valor $p(x)$, asociado al vértice seleccionado, se deja fijo para las demás iteraciones.

- En las iteraciones siguientes, para cada vértice v_j que todavía no forme parte de la columna V (es decir, que no forma parte del árbol) se actualiza $p(v_j)$ en caso de que v_j sea adyacente a x , el cual pasa a ser igual a:

$$\min \{p(v_j), p(v_i) + p(\{x, v_i\})\},$$

siendo v_i el vértice seleccionado en el paso anterior, obteniéndose de esa forma el peso del camino mínimo que va desde v_1 hasta v_j .

El valor mínimo del conjunto $\{p(v_j)\}$ formado por nodos que no forman parte de la columna V en el momento actual, proporciona un nuevo vértice, al que llamamos x . Entonces añadimos el vértice x en la columna V y a la columna E se le añade la arista $\{a, x\}$, siendo $a = x$ si el mínimo se ha dado en el segundo argumento y $a = v_{k-1}$, siendo k la primera fila a partir de la cual los valores asociados al vértice x son constantes hasta el paso actual. Este último caso se produce cuando la longitud del camino asociado a x , obtenido en iteraciones anteriores, es menor que la resultante de conectar el camino que produce el algoritmo con dicho vértice, lo cual obliga a deshacer algunos pasos.

- El proceso continúa hasta añadir v_n en la columna V.

Ejemplo:

Consideremos el grafo de seis vértices con la representación gráfica que se muestra a continuación, con los correspondientes valores numéricos de los pesos indicados en cada una de sus aristas:

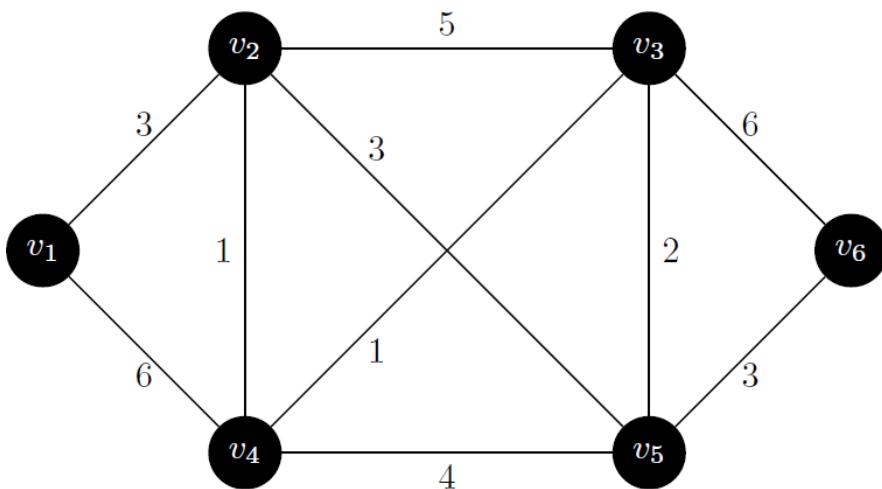


Figura 36. Grafo sobre el que trabajaremos en el ejemplo. Fuente: elaboración propia.

Queremos encontrar el camino más corto entre v_1 y v_6 , es decir, el camino formado por la secuencia de aristas cuya suma de pesos sea la mínima posible. Para ello, hallaremos el árbol que proporciona el algoritmo de Dijkstra y, a continuación, obtendremos su único camino entre v_1 y v_6 , que proporcionará en última instancia el camino de menor longitud.

Seguimos los pasos siguientes:

- Definimos $p(v_i) = 0$ y, para $2 \leq j \leq 6$, definimos $p(v_j)$ de forma que sea el peso de la arista que lo une con v_1 (en caso de existir) o bien ∞ en caso contrario. Así pues, obtenemos la tabla siguiente asociada a la primera iteración:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞		

Ahora, seleccionamos el vértice (excluyendo v_1) tal que el peso asignado es mínimo. Vemos que en este caso el resultado es v_2 , con $p(v_2) = 3$, el cual es el valor mínimo que aparece en la tabla (sin contar v_1), por lo que fijamos ese valor (no lo variaremos en futuras iteraciones). Anotamos dicho vértice en la columna **V** y la correspondiente arista que lo une con v en **E**, es decir, $\{v_1, v_2\}$, por lo que la tabla queda de la forma siguiente:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$

Por tanto, gráficamente el proceso algorítmico de construcción del árbol, señalando los correspondientes vértices y aristas con color cian, se encuentra como se muestra a continuación.

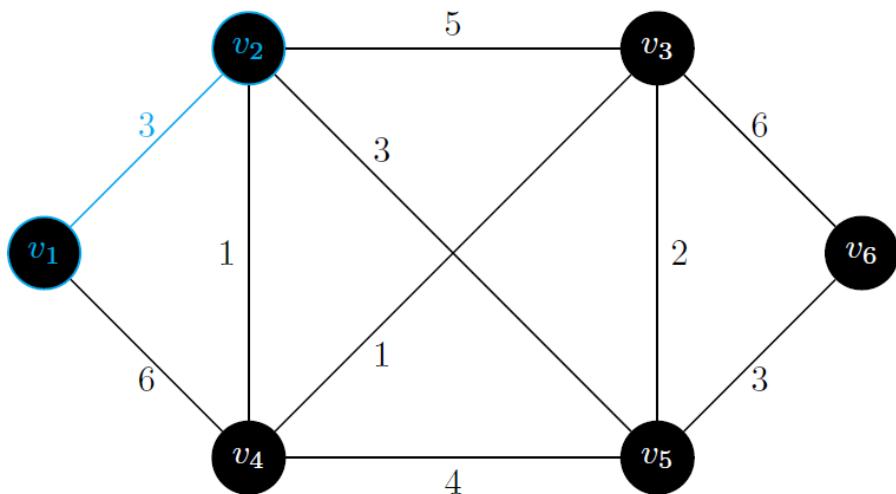


Figura 37. Estado del proceso algorítmico tras la primera iteración. Fuente: elaboración propia.

2. Tomamos en consideración ahora el vértice anteriormente escogido, es decir, v_2 , y, en caso de que v_j sea adyacente a v_2 , $v_j \neq v_1$, $v_j \neq v_2$, cambiamos ahora los correspondientes valores de $p(v_j)$, por el valor:

$$\min \{p(v_j), p(v_2) + p(\{v_2, v_j\})\}$$

Los vértices que cumplen esta condición son v_3 , v_4 y v_5 , siendo:

$$\min \{p(v_3), p(v_2) + p(\{v_2, v_3\})\} = \min \{\infty, 3 + 5\} = 8,$$

$$\min \{p(v_4), p(v_2) + p(\{v_2, v_4\})\} = \min \{6, 3 + 1\} = 4,$$

$$\min \{p(v_5), p(v_2) + p(\{v_2, v_5\})\} = \min \{\infty, 3 + 3\} = 6,$$

Así pues, la tabla actualizada con la segunda iteración queda como sigue:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞		

Ahora, entre los valores presentes en la segunda iteración, seleccionamos el menor de todos, que en este caso se corresponde al vértice v_4 con el valor 4. Así pues, fijamos este valor, que ya no variará en las próximas iteraciones, anotamos en la columna V el vértice v_4 y, como el mínimo se ha dado en el segundo argumento, en la columna E la arista de menor peso (longitud) que lo une con el camino anterior, es decir $\{v_2, v_4\}$. Por tanto, la tabla actualizada es la siguiente:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$

Con lo anterior, el estado visual de la creación del árbol se corresponde con la representación gráfica siguiente:

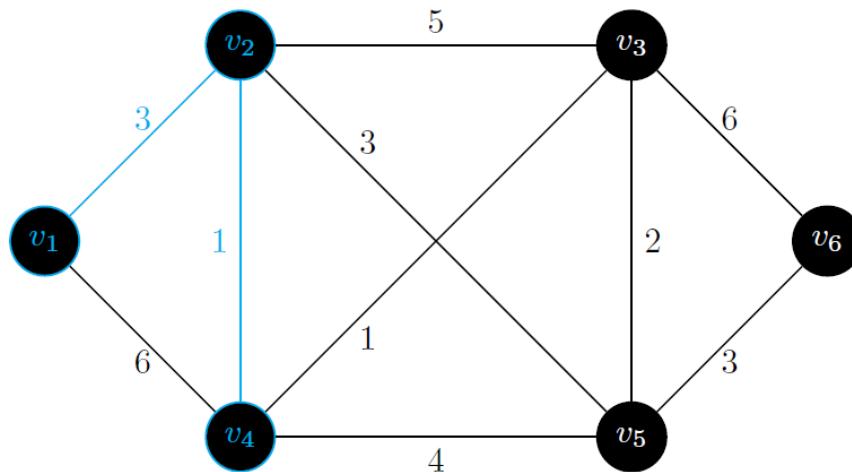


Figura 38. Estado del proceso algorítmico tras la segunda iteración. Fuente: elaboración propia.

3. El vértice previamente elegido es v . Ahora, en caso de que v_j sea adyacente a v_4 , $v_j \neq v_1$, $v_j \neq v_2$, $v_j \neq v_4$ (o, lo que es lo mismo, $v_2 \in \{v_3, v_5, v_6\}$), sustituimos $p(v_i)$ por:

$$\min \{p(v_j), p(v_4) + p(\{v_4, v_j\})\}$$

En este caso, los vértices considerados que verifican ser adyacentes con v_4 son v_3 y v_5 , por lo que se tiene:

$$\begin{aligned} \min \{p(v_3), p(v_4) + p(\{v_4, v_3\})\} &= \min \{8, 4+1\} = 5, \\ \min \{p(v_5), p(v_4) + p(\{v_4, v_5\})\} &= \min \{6, 4+4\} = 6, \end{aligned}$$

Por tanto, la tabla con los datos actualizados tiene el aspecto siguiente:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$
3			5		6	∞		

En este caso, entre los valores que aparecen en la segunda iteración, el más bajo es el 5, que se encuentra en la columna correspondiente al peso $p(v_3)$. Así pues, el vértice seleccionado esta vez es v_3 , por lo que lo añadimos en la columna V, así como la arista que lo une con el camino anteriormente elegido (pues el mínimo se ha dado en el segundo argumento), es decir, $\{v_4, v_3\}$, ya que es la que tiene un peso menor. Por último, congelamos el valor $p(v_3)$, que ya no variará en las iteraciones siguientes. Por tanto, la tabla actualizada tiene el aspecto siguiente:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$
3			5		6	∞	v_3	$\{v_4, v_3\}$

Gráficamente:

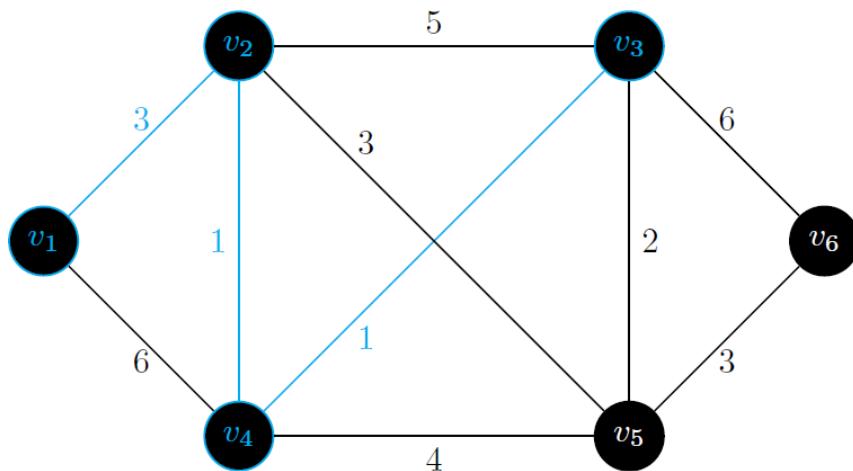


Figura 39. Estado del proceso algorítmico tras la tercera iteración. Fuente: elaboración propia.

4. El vértice anteriormente seleccionado es v_3 , siendo v_1 , v_2 y v_4 los que le preceden, respectivamente. Así pues, entre los nodos restantes, $v_j \in \{v_5, v_6\}$, debemos modificar el valor $p(v_j)$ en caso de que sean adyacentes a v_3 por

$$\min \{p(v_j), p(v_3) + p(\{v_3, v_j\})\}$$

En este caso, ambos vértices son adyacentes a v_3 , por lo que se tienen las siguientes modificaciones:

$$\begin{aligned} \min \{p(v_5), p(v_3) + p(\{v_3, v_5\})\} &= \min \{6, 5+2\} = 6, \\ \min \{p(v_6), p(v_3) + p(\{v_3, v_6\})\} &= \min \{\infty, 5+6\} = 11, \end{aligned}$$

En consecuencia, la tabla actualizada con los datos de la cuarta iteración es la siguiente:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$
3			5		6	∞	v_3	$\{v_4, v_3\}$
4					6	11		

Se observa por tanto que el vértice con menor peso es en este caso v_5 , por lo que lo añadimos en la columna V y, en lo que respecta a la arista, observamos que en este caso el mínimo se ha dado en el primer argumento, por lo que, observando la tabla, vemos que el valor $p(v_5)$ se repite hasta la segunda iteración, donde el nodo de partida era v_2 , por lo que se añade la arista $\{v_2, v_5\}$.

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$
3			5		6	∞	v_3	$\{v_4, v_3\}$
4					6	11	v_5	$\{v_2, v_5\}$

Gráficamente:

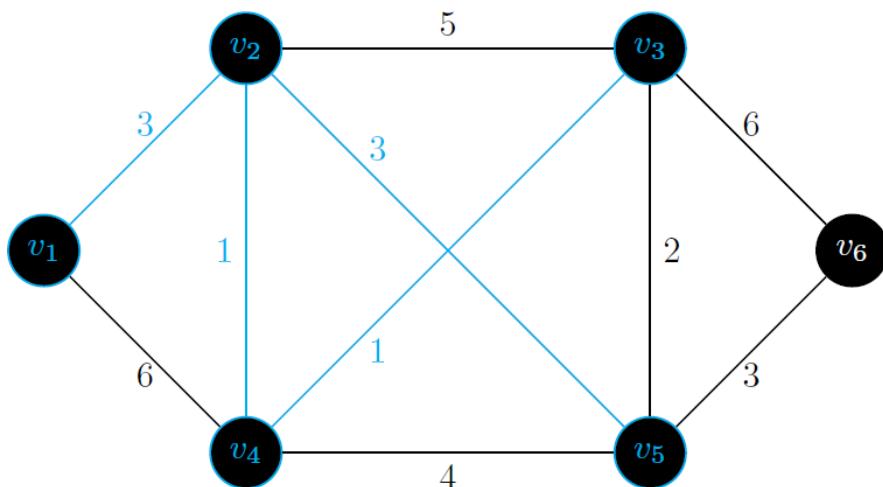


Figura 40. Estado del proceso algorítmico tras la cuarta iteración. Fuente: elaboración propia.

5. Partimos del vértice seleccionado en la iteración anterior, v_5 . Ahora, el único vértice que nos queda por elegir es v_6 , que es adyacente a v_5 , por lo que calculamos

$$\min \{ p(v_j), p(v_5) + p(\{v_5, v_j\}) \}$$

Como en este caso solo debemos realizar la operación para $v = 6$, se tiene el resultado siguiente:

$$\min \{ p(v_6), p(v_5) + p(\{v_5, v_6\}) \} = \min \{ 11, 6 + 3 \} = 9,$$

Por tanto, la tabla con los datos actualizados es:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$
3			5		6	∞	v_3	$\{v_4, v_3\}$
4					6	11	v_5	$\{v_2, v_5\}$
5						9		

Como 9 es el único valor, asociado al vértice v_6 , lo seleccionamos para introducirlo en la columna V. Por otra parte, el mínimo se ha dado en el segundo argumento, lo cual indica que la conexión con el camino más corto de v_5 es menor que $p(v_6)$, por lo que añadimos a la columna E la arista $\{v_5, v_6\}$. Finalmente, fijamos el valor $p(v_6)$ y la tabla, por tanto, queda con el aspecto siguiente:

Iteración	$p(v_1)$	$p(v_2)$	$p(v_3)$	$p(v_4)$	$p(v_5)$	$p(v_6)$	V	E
1	0	3	∞	6	∞	∞	v_2	$\{v_1, v_2\}$
2			8	4	6	∞	v_4	$\{v_2, v_4\}$
3			5		6	∞	v_3	$\{v_4, v_3\}$
4					6	11	v_5	$\{v_2, v_5\}$
5						9	v_6	$\{v_5, v_6\}$

Gráficamente:

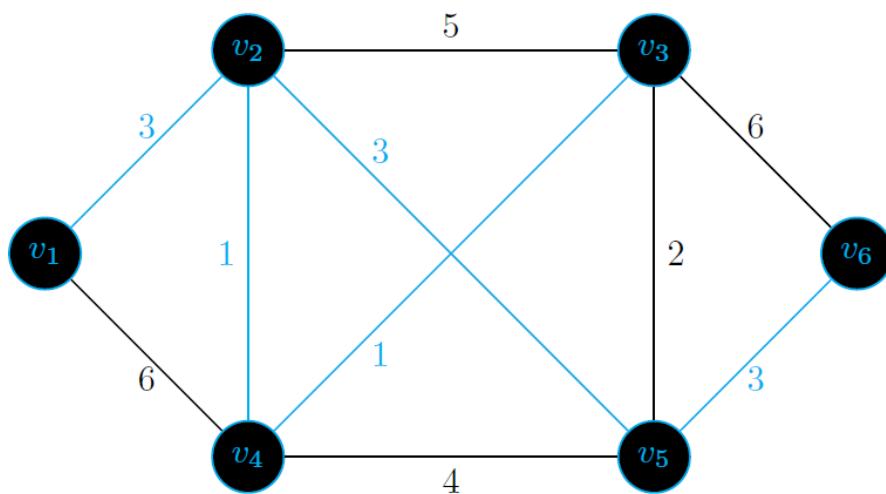


Figura 41. Estado del proceso algorítmico tras la quinta iteración. Fuente: elaboración propia.

Se concluye, por tanto, que el camino más corto tiene longitud 9 y, siguiendo la tabla de abajo a arriba, el camino formado en última instancia está formado por los vértices $\{v_1, v_2\}$, $\{v_2, v_5\}$ y $\{v_5, v_6\}$.

Se muestra en la ilustración siguiente el camino final, es decir, el camino más corto entre v_1 y v_6 resaltado en color naranja.

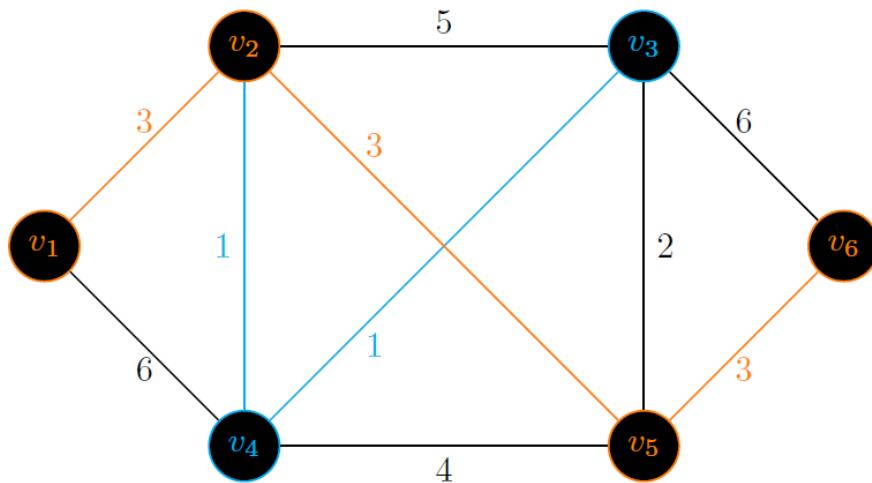


Figura 42. Camino resultante (destacado en naranja). Fuente: elaboración propia.

Resumen

Para que una máquina aprenda y muestre tener una inteligencia como tal, es necesario que se la haya programado previamente para tomar decisiones en base a verdades y a un proceso de toma de decisiones. En este sentido, la lógica es la ciencia matemática que ayudará al profesional a programar las máquinas de tal manera que no solo tome las decisiones en base a unos estímulos externos, sino que también aprenda por sí misma, una vez comience a recibir inputs del entorno.

Dentro de la lógica como ciencia matemática, se dan diferentes formas de acercamiento. En primer lugar, se ha hablado de la lógica proposicional, que establece la relación y conexión entre diferentes enunciados. Esta relación se ha estudiado concretamente a través del álgebra de Boole. Esta será de utilidad en el futuro para simplificar afirmaciones que a priori son complejas. Se ha hablado en este caso de los enunciados lógicos o proposiciones lógicas, como cualquier afirmación a la que se le puede atribuir un valor de verdad y que, por tanto, está asociado al concepto de verdadero o falso.

Las funciones booleanas, por su parte, se han definido y trabajado ya que son útiles en cuanto que permiten construir puertas lógicas a través de las cuales se construyen los circuitos como tales. Las puertas lógicas son aquellas operaciones que, siendo programadas después en la máquina, permiten predecir el comportamiento de esta a través de operaciones binarias. En este apartado se han hablado de la Puerta OR, Puerta AND, Puerta XOR, Puerta NOT, Puerta NOR, Puerta NAND y, finalmente, Puerta XNOR.

La lógica de predicados, por otro lado, ha sido definida como aquella teoría que permite formalizar y estructurar el pensamiento lógico. Aunque esta teoría no permite cubrir la totalidad de situaciones que pueden darse en el día a día, sí resulta útil para la IA en cuanto que permite resolver parte de ella. En este apartado se ha hablado de los conceptos de predicados y universos, los cuantificadores y las operaciones realizadas para el cálculo de predicados.

La lógica inductiva se ha definido como aquella en la que, partiendo de premisas concretas, se llega a una conclusión general y amplia que no tendrá la garantía per se de ser cierta. En este apartado, se ha trabajado en cómo definir las hipótesis que permitirán llegar a una conclusión concreta, a través de la inducción predictiva y descriptiva.

La complejidad computacional es la magnitud de la complejidad de un algoritmo que da solución a un problema decidable o computable. Un problema como tal tiene solución si se puede resolver con un sí o no. En este punto se han trabajado también los tipos de complejidad de un problema.

Por último, se ha revisado la teoría de grafos, como herramienta diseñada para resolver de manera elegante y efectiva un problema. En este apartado se han mostrado los aspectos más relevantes de esta teoría y su aplicación a la IA: grafo, subgrafo, grafo simple, grafo dirigido y la matriz de adyacencia. Además, se han trabajado los conceptos de caminos y conexiones, así como los árboles.



Capítulo 2

Álgebra

Introducción

El álgebra es aquella rama de las matemáticas en la que se utilizan elementos numéricos y letras y se realizan operaciones entre ellas. De forma sencilla, el álgebra como tal es utilizada en Inteligencia Artificial para poder trabajar y, por tanto, predecir situaciones en las que se tiene un gran volumen de datos a analizar y tener en cuenta.

De esta manera, durante el presente capítulo se realizará un acercamiento a la teoría de conjuntos, así como a la forma en la que se establecen las relaciones y aplicaciones de los elementos que los forman.

Se hablará también de los espacios vectoriales, la factorización de matrices y el cálculo tensorial.

Todo ello siempre desde el punto de vista de que el álgebra permite y facilita el trabajo de las futuras máquinas y su posterior aprendizaje de las situaciones en las que se encuentren.

Objetivos

- Recordar los elementos básicos que forman la rama del álgebra y su aplicación en la Inteligencia Artificial.
- Realizar una relación entre conjuntos, para poder trabajar con grandes volúmenes de datos.
- Aplicar los conocimientos sobre álgebra lineal para el manejo de grandes volúmenes de datos, factorización de matrices y cálculo tensorial.

2.1. Teoría de conjuntos

Los conceptos de **conjunto** y **elemento** (de un conjunto) no disponen de una definición formal desde el punto de vista matemático, si bien la intuición permite hacerse una idea de su significado como entes abstractos. Así pues, un conjunto puede entenderse como una colección de objetos que se encuentran en una misma categoría, en calidad de elementos de este, como entes **pertenecientes** a dicho conjunto. Estos pueden ser finitos o infinitos en función del número de elementos que tengan y pueden representarse entre llaves listando cada uno de sus elementos separados por una coma, como veremos en los ejemplos que se muestran a continuación:

- Días de la semana:
 $\{\text{lunes, martes, miércoles, jueves, viernes, sábado, domingo}\}$
- Posibles valores obtenidos al lanzar un dado de parchís:
 $\{1, 2, 3, 4, 5, 6\}$
- Las tres primeras letras del abecedario en minúsculas:
 $\{a, b, c\}$
- Números naturales pares:
 $\{2, 4, 6, 8, 10, 12, \dots\}$
- Números primos menores que 20:
 $\{2, 3, 5, 7, 11, 13, 17, 19\}$

Como hemos podido observar, el penúltimo ejemplo consiste en un conjunto infinito, motivo por el cual se ha dejado indicado con puntos suspensivos que los elementos continúan indefinidamente en la progresión indicada. Existen ciertos tipos de conjuntos infinitos, como el que nos acontece, que pueden expresarse de una forma más elegante desde el punto de vista matemático, haciendo un uso adecuado de la notación. Así pues, el conjunto de los números naturales pares podría representarse como sigue:

$$\{2n / n \in \mathbb{N}\}.$$

La expresión anterior puede leerse de izquierda a derecha como “el conjunto de los valores de la forma $2n$ tales que n pertenece a los números naturales”, que corresponde, efectivamente, al conjunto formado por los números pares, los cuales son todos los múltiplos de 2.

Notas:

- Cada uno de los elementos de un conjunto solo debe aparecer una única vez; es decir, no pueden repetirse elementos. Por ejemplo, es incorrecto expresar el conjunto formado por las tres primeras letras del abecedario como $\{a, b, c, c, c, a\}$, ya que tanto el elemento a como el elemento c aparecen repetidos al menos una vez.
- De entre todos los conjuntos, cabe destacar un caso especial, que es el **conjunto vacío**, formado por cero elementos, y que se denota como \emptyset .

Definición:

El cardinal de un conjunto A es el número de elementos que tiene, y se denota por $|A|$.

- Si el conjunto es finito y tiene n elementos, decimos que $|A|=n$.
- Si, por el contrario, tiene infinitos elementos, se escribe $|A|=\infty$.

2.1.1. Operaciones entre conjuntos

Además de haber formas de relacionar conjuntos, también existen operaciones entre ellos que son ampliamente utilizadas en diferentes campos de las matemáticas. Veremos a continuación las operaciones más básicas entre ellos.

Definición:

La **intersección** de dos conjuntos A,B es el conjunto formado por todos los elementos comunes a ambos conjuntos y se denota por $A \cap B$, es decir, los elementos que pertenecen *simultáneamente* a A y B. Formalmente, puede definirse como sigue:

$$A \cap B = \{x / x \in A \wedge x \in B\}.$$

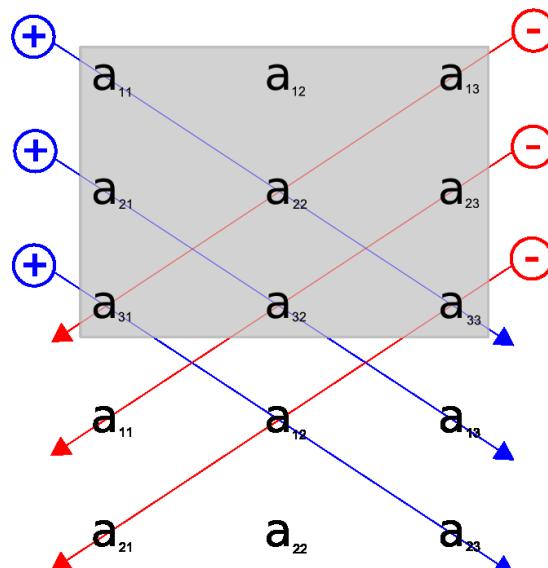


Figura 43. Ejemplo en el que se ilustra gráficamente la noción de intersección entre dos conjuntos.


Ejemplo

1. Sean $A=\{-1, 3, 5, 2, 6, 9\}$ y $B=\{-1, 0, 4, 3, 7, 9, 10\}$. Para calcular su intersección, observamos qué elementos tienen en común, que en este caso son los valores $-1, 3$ y 9 . Por tanto, se tiene que $A \cap B = \{-1, 3, 9\}$
2. Consideremos ahora $A=\{a, b, c\}$ y $B=\{d, e, f\}$. En este caso, observamos que no presentan ningún elemento en común. Cuando esto ocurre, decimos que la intersección es **vacía** e igual al conjunto vacío, definido anteriormente como el conjunto que no está formado por ningún elemento: $A \cap B = \emptyset$. En este caso, se dice que A y B son **disjuntos**.

Definición:

La unión entre dos conjuntos A, B es el conjunto formado tanto por los elementos de A como por los elementos de B y se denota por $A \cup B$, es decir, los elementos que pertenecen a A o bien a B . Su expresión formal es la siguiente:

$$A \cup B = \{x / x \in A \vee x \in B\}.$$

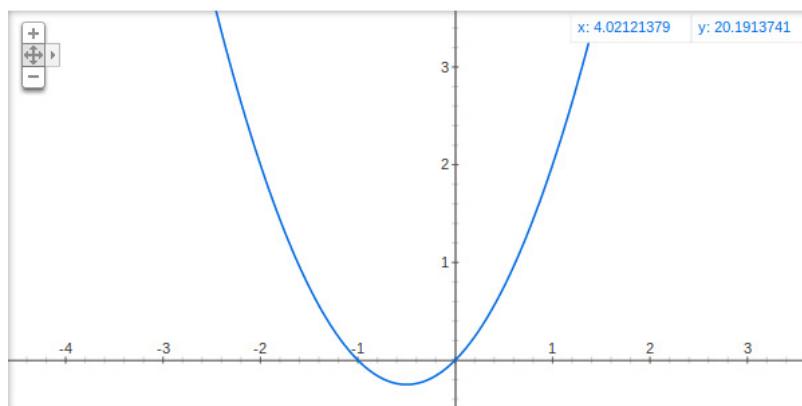


Figura 44. Ejemplo en el que se ilustra gráficamente la noción de unión entre dos conjuntos.


Ejemplo

1. Sean $A=\{a, b, c, e\}$ y $B=\{c, e, f\}$. Su unión está formada por los elementos tanto de A como de B , luego $A \cup B = \{a, b, c, e, f\}$. Obsérvese que tanto el elemento c como el elemento e son comunes a ambos conjuntos, pero no se han repetido en el resultado, ya que, tal y como se ha indicado anteriormente, es incorrecto repetir elementos en los conjuntos.
2. Se consideran $A=\{1, 3, 5\}$ y $B=\{2, 4, 6\}$. En este caso los conjuntos son disjuntos entre sí (no tienen ningún elemento en común) y su unión es $A \cup B = \{1, 2, 3, 4, 5, 6\}$.

Definición:

El **complementario** de un conjunto A sobre otro conjunto B es el conjunto formado por todos los elementos de B que no pertenecen a A y se denota por $B \setminus A$. Formalmente se expresa como sigue:

$$B \setminus A = \{x / x \in B \wedge x \notin A\}.$$

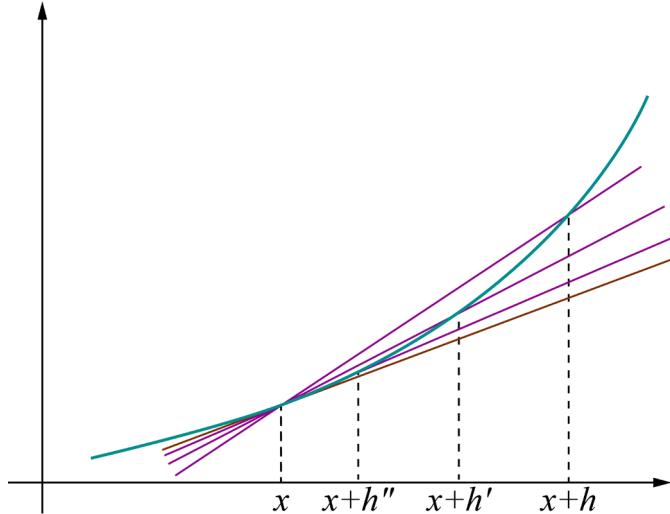


Figura 45. Ejemplo en el que se ilustra gráficamente la noción de complementario.

Nota:

En el caso en que se trabaje con un conjunto X que se toma como el total, es decir, que contiene todos los elementos posibles con los que se trabaja, podemos hablar del complementario de A (sin especificar sobre qué conjunto), entendiendo que se realiza sobre el conjunto total X. En este caso, se denota $A^c = X \setminus A$.

**Ejemplo**

1. Sean $A = \{a, b, c, f\}$ y $B = \{b, c, e, f, g, j\}$. Entonces $B \setminus A$ consiste en quitar a B todos los elementos que estén en A, en este caso, b, c y f, por lo que $B \setminus A = \{e, g, j\}$.
2. Definimos ahora $A = \{a, b, c\}$ y $B = \{d, e, f\}$. En este caso, B no tiene ningún elemento que forme parte de A, por lo que el complementario de A con respecto a B continúa siendo B en tanto que no se elimina ninguno de sus elementos, esto es, $B \setminus A = \{d, e, f\} = B$.
3. Consideremos $A = \{a, b, c, d\}$ y $B = \{b, c\}$. Ahora se observa que todos los elementos de B forman parte de A, luego el resultado de realizar el complementario de A sobre B consiste en eliminar todos los elementos de este último, con lo cual se obtiene el conjunto vacío: $B \setminus A = \emptyset$.
4. Sea $A = \{2, 4, 6, 8, \dots\} = \{2n / n \in \mathbb{N}\}$ el conjunto formado por los números pares sobre el conjunto total de los números naturales, \mathbb{N} . Entonces el complementario de A es el conjunto formado por los números impares, es decir, $A^c = \{1, 3, 5, 7, 9, \dots\} = \{2n+1 | n \in \mathbb{N}\}$.

Propiedades:

Sean A, B, C conjuntos cualesquiera. Se cumplen las siguientes relaciones:

1. $(A \cup B) \cup C = A \cup (B \cup C)$.
2. $(A \cap B) \cap C = A \cap (B \cap C)$.
3. $A \cup B = B \cup A$.
4. $A \cap B = B \cap A$.
5. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
6. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
7. $(A \cup B)^c = A^c \cap B^c$.
8. $(A \cap B)^c = A^c \cup B^c$.

2.1.2. Partes de un conjunto

Es preciso recordar que cuando se trabaja con conjuntos y sus elementos, estos últimos pueden ser de todo tipo, sin limitación en el nivel abstracción que los determina: números, letras, días de la semana, colores, figuras..., o incluso conjuntos en sí mismos. Es decir, podemos considerar conjuntos cuyos elementos sean a su vez conjuntos que contengan uno o varios elementos y así sucesivamente, de manera que se crea una relación de jerarquía.

**Ejemplo**

Consideremos el conjunto $A = \{(a, b), c\}$. Obsérvese que este conjunto tiene dos elementos, uno de los cuales es, a su vez, otro conjunto, $\{a, b\} \in A$, y el otro es $c \in A$. Al mismo tiempo, tenemos el conjunto $\{a, b\}$ que, como decíamos, tiene el rol de elemento de A, y que, por su parte, está formado por dos elementos, $a \in \{a, b\}$ y $b \in \{a, b\}$. Esta relación no implica que los elementos a y b pertenezcan a A; de hecho, en este caso es falso, ya que A solo tiene dos elementos: $\{a, b\}$ y c, y, como podemos observar, ninguno de los dos son ni a ni b.

A partir de cada conjunto, puede construirse otro conjunto de interés que se define como sigue.

Definición:

Sea A un conjunto. El conjunto de las **partes de A**, denotado por $P(A)$, consiste en el conjunto formado por todos los subconjuntos posibles de A; es decir, es el conjunto cuyos elementos son todos los subconjuntos de A:

$$P(A) = \{B / B \subseteq A\}.$$

Ejemplo:

Tomemos el conjunto $A = \{1, 2, 3\}$ y calculemos $P(A)$. Para ello, vamos a calcular todos los subconjuntos posibles de A separándolos por número de elementos (cardinal).

Cardinal 0: \emptyset .

Cardinal 1: {1}, {2}, {3}.

Cardinal 2: {1,2}, {1, 3}, {2,3}.

Cardinal 3: {1, 2, 3}.

Así pues, el conjunto partes de A está formado por 8 elementos (1 de cero elementos, 3 de un elemento, 3 de dos elementos y 1 de tres elementos) y es:

$$P(A)=\{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1, 3\}, \{2,3\}, \{1, 2, 3\}\}$$

Comúnmente, puede generalizarse la siguiente propiedad, que relaciona el cardinal de un conjunto finito con el de sus partes.

Teorema:

Sea A un conjunto finito. Si $|A|=n$, entonces $|P(A)|=2^n$.

2.2. Relaciones y aplicaciones

Dados dos conjuntos A, B, pueden establecerse relaciones entre los elementos de ambos conjuntos. Para formalizar esta noción, en primer lugar necesitamos establecer la noción de **par ordenado**.

Definición:

Sean A, B dos conjuntos, $a \in A$ y $b \in B$ elementos de A y B, respectivamente. Podemos unificar ambos elementos siguiendo un orden y formar un nuevo elemento, denotado por (a, b) , que recibe el nombre de **par ordenado**, cuya primera componente es a y con segunda componente b.

Esta noción permite establecer, a su vez, la definición de **producto cartesiano**:

Definición:

Sean A, B dos conjuntos. El producto cartesiano de A y B, denotado por $A \times B$, consta del conjunto formado por todos los pares ordenados que pueden formarse con elementos de A en primer lugar y elementos de B en segundo lugar. Formalmente:

$$A \times B = \{(a,b) \mid a \in A \wedge b \in B\}.$$

Ejemplo

1. Sea $A=\{1, 2, 3\}$ y $B=\{c, d\}$. El producto cartesiano de A y B está formado por todas las combinaciones posibles de pares ordenados con primer elemento en A y segundo elemento en B, luego:

$$A \times B = \{(1,c), (1,d), (2,c), (2,d), (3,c), (3,d)\}.$$

2. Consideremos esta vez $A = \mathbb{R}$ y $B = \mathbb{R}$. Entonces el producto cartesiano de ambos conjuntos es el plano real, ya que:

$$A \times B = \mathbb{R} \times \mathbb{R} = \{(x,y) \mid x \in \mathbb{R} \wedge y \in \mathbb{R}\} = \mathbb{R}^2.$$

Un cálculo sencillo permite determinar el número de elementos de un producto cartesiano en función del número de elementos de los dos conjuntos que lo forman.

Tras estas definiciones preliminares, ya estamos en condiciones de introducir el concepto de relación. Dados dos conjuntos A y B, podemos relacionar diferentes elementos de A con diferentes elementos de B mediante una regla específica. Al final, esto puede identificarse como un conjunto concreto de pares de elementos (a, b), con $a \in A$ y $b \in B$ que cumplen la relación, es decir, que verifiquen que a esté relacionado con b. Así pues, una relación determina en realidad un subconjunto de $A \times B$, que se corresponde con la definición formal que introduciremos de relación.

2.2.1. Relaciones entre conjuntos

Definición:

Sean A, B conjuntos cualesquiera. Una relación R de A en B es un subconjunto de $A \times B$, cuyos pares representan los elementos que están relacionados entre sí. Si $a \in A$ y $b \in B$ están relacionados entre sí, se denota aRb . Por tanto, puede escribirse:

$$R = \{(a,b) \in A \times B \mid aRb\} \subseteq A \times B.$$

Ejemplo

1. Tomamos $A = \{1, 2, 3, 4, 5\}$ y $B = \{-2, 0, 1, -3\}$ y, fijando $a \in A$ y $b \in B$, establecemos la relación aRb si a es par y b es negativo. Así pues, se tiene:
- $$R = \{(2, -2), (2, -3), (4, -2), (4, -3)\}.$$
2. Siguiendo con los mismos conjuntos de antes, tomamos ahora la relación aRb si su suma, $a+b$, es múltiplo de 3. Sumando cada elemento de A con cada elemento de B y observando en qué casos el resultado es múltiplo de 3, se obtiene:
- $$R = \{(2, -2), (2, 1), (3, 0), (3, -3), (5, -2), (5, 1)\}.$$

2.2.2. Relaciones de equivalencia

Existe un caso especial de relaciones particularmente interesantes por sus propiedades y utilidad para construir nuevas estructuras, que son las llamadas *relaciones de equivalencia*, las cuales vamos a tratar a continuación.

Definición:

Una relación R de un conjunto A sobre sí mismo se dice que es de **equivalencia** si cumple las tres propiedades siguientes:

1. **Reflexiva:** $\forall a \in A, aRa$.
2. **Simétrica:** $\forall a, b \in A, aRb \rightarrow bRa$.
3. **Transitiva:** $\forall a, b, c \in A, aRb \wedge bRc \rightarrow aRc$.



Ejemplo

1. Sea A un conjunto cualquiera. El ejemplo trivial de relación de equivalencia es la igualdad en sí misma: dados, $a,b \in A$, $a \mathrel{R} b$ si $a=b$. Comprobemos que efectivamente es una relación de equivalencia:
 - a. Reflexiva: sea $a \in A$. Como $a=a$, se tiene por definición que aRa .
 - b. Simétrica: sean $a,b \in A$, y supongamos que $a \mathrel{R} b$. Por definición, se tiene que $a=b$, que puede escribirse como $b=a$, lo cual, nuevamente por definición, se traduce en bRa .
 - c. Transitiva: sean $a, b, c \in A$, con $a \mathrel{R} b$ y $b \mathrel{R} c$. Veamos que $a \mathrel{R} c$. En efecto, de aRb se tiene por definición que $a=b$ y de bRc se tiene $b=c$. Por tanto, se concluye que $a=c$, por lo que por definición se obtiene $a \mathrel{R} c$.
2. Consideraremos ahora el conjunto de los números enteros, \mathbb{Z} , y establezcamos la siguiente relación: dados $a,b \in \mathbb{Z}$, $a \mathrel{R} b$ si $\exists m \in \mathbb{Z} / b-a=3m$, es decir, si su diferencia es un múltiplo de 3. Veamos que es una relación de equivalencia:



Ejemplo

- a. Reflexiva: sea $a \in \mathbb{Z}$. Entonces $a-a=0=3 \cdot 0$ (obviamente, 0 es un múltiplo de 3), luego aRa .
- b. Simétrica: sean $a,b \in \mathbb{Z}$ y supongamos que $a \mathrel{R} b$. Eso implica que $\exists m \in \mathbb{Z}$ tal que $b-a=3m$. Ahora, para probar que bRa , tenemos que encontrar $m' \in \mathbb{Z}$ tal que $a-b=-3m'$, pero esto es inmediato, ya que como $b-a=3m$, se tiene que $a-b=-3m=3 \cdot (-m)$, por lo que tomando $m'=-m$ se tiene el resultado deseado y se demuestra que bRa .
- c. Transitiva: sean $a,b,c \in \mathbb{Z}$, con $a \mathrel{R} b$ y $b \mathrel{R} c$. Entonces se tiene por una parte que $b-a=3m$ y por otra parte que $c-b=3m'$. Para probar que $a \mathrel{R} c$, tenemos que encontrar $m'' \in \mathbb{Z}$ tal que $c-a=3m''$. Esto es inmediato utilizando las igualdades anteriores, ya que $c-a=c-b+b-a=3m'-3m=3(m'-m)$. Así pues, basta con tomar $m''=m'-m$, siendo este el valor que buscábamos, para concluir que $a \mathrel{R} c$.

Las relaciones de equivalencia determinan una serie de elementos que están relacionados entre sí y los dividen en grupos según la naturaleza de la relación en cuestión. Por una parte, en el primer ejemplo podemos ver que, dado un elemento cualquiera, este solo está relacionado con sí mismo y con ninguno más, por lo que cada elemento queda aislado dentro de la relación en cuestión. Por otra parte, el segundo ejemplo da más de sí, de modo que el 0 está relacionado con el 3, 6, 9, ... y así sucesivamente, así como con el -3, -6, -9, ...; el 1 con el 4, 7, 10, ... y con el -2, -5, -8, ...; el 2 con el 5, 8, 11, ... y con el -1, -4, -7, ...; etc.

Veremos a continuación que esta división está bien estructurada cuando se trata de relaciones de equivalencia, en el mismo sentido que los grupos de elementos relacionados entre sí están bien determinados y separados de los demás. Dichos grupos reciben el nombre de clase de equivalencia, que se define como sigue.

Definición:

Sea A un conjunto cualquiera, \mathfrak{R} una relación de equivalencia sobre A y $a \in A$. Se define la **clase de equivalencia** del elemento $a \in A$ asociada a la relación R como sigue:

$$[a]_{\mathfrak{R}} = \{b \in A / a \mathfrak{R} b\}.$$

Nota:

$\forall a \in A, [a]_{\mathfrak{R}} \neq \emptyset$, ya que es evidente que por ejemplo $a \in [a]_{\mathfrak{R}}$, puesto que aRa . Al elemento que aparece entre los corchetes, en este caso a , suele llamársele el **representante de la clase**.

Teorema:

Sea \mathfrak{R} una relación de equivalencia sobre un conjunto A y $a, b \in A$. Las afirmaciones siguientes son equivalentes:

1. $a \mathfrak{R} b$.
2. $[a]_{\mathfrak{R}} = [b]_{\mathfrak{R}}$.
3. $[a]_{\mathfrak{R}} \cap [b]_{\mathfrak{R}} \neq \emptyset$.

Obsérvese que la propiedad anterior nos indica que solo puede haber dos posibilidades cuando se toman las clases de equivalencia de dos elementos cualesquiera: que ambos elementos estén relacionados, en cuyo caso es equivalente a que sus clases de equivalencia sean los mismos conjuntos; o bien que no estén relacionados, lo cual implica que no solo sus clases son diferentes como conjuntos, sino que como tales no tienen nada que ver, ya que son disjuntos, es decir, no tienen ningún elemento en común.

**Ejemplo**

Recordemos los ejemplos anteriores para ver qué clases de equivalencia se obtienen a partir de cada una de las relaciones de equivalencia.

1. Dado el conjunto A , tomamos la relación de equivalencia aRb si $a=b$. Fijado un elemento $a \in A$, se tiene $[a]_{\mathfrak{R}} = \{b \in A / a \mathfrak{R} b\} = \{b \in A / a=b\} = \{a\}$. Así pues, en este caso toda clase de equivalencia asociada a un elemento a cualquiera está formada únicamente por ese mismo elemento; es decir, esta relación de equivalencia induce clases de equivalencia lo más pequeñas posible, habiendo tantas como elementos de A .
2. Tomemos ahora la relación de equivalencia sobre \mathbb{Z} $a \mathfrak{R} b$ si $b-a$ es múltiplo de 3. Por ejemplo, al tomar el elemento 0, se observa que este está relacionado con todos los múltiplos de 3. Por otra parte, al tomar el elemento 1, está relacionado con los números de la forma $3m+1$, puesto que su diferencia en cualquier caso es un múltiplo de 3. Finalmente, al tomar el elemento 2, este está relacionado con todos los números de la forma $3m+2$, ya que su diferencia en cualquiera de los casos es también un múltiplo de 3. En definitiva, hay tres clases de equivalencia, que vienen determinadas por el resto obtenido al dividir entre 3, con los números 0, 1 y 2 como representantes de cada clase:

- a. $[0]_{\mathfrak{R}} = \{\dots, -9, -6, -3, 0, 3, 6, 9, \dots\}$.
- b. $[1]_{\mathfrak{R}} = \{\dots, -8, -5, -2, 1, 4, 7, 10, \dots\}$.
- c. $[2]_{\mathfrak{R}} = \{\dots, -7, -4, -1, 2, 5, 8, 11, \dots\}$.

Por último, nótese que podría haberse elegido cualquier otro representante para cada clase, siempre y cuando se hubiese tratado de un elemento de esta. Por ejemplo, $[0]_{\mathfrak{R}} = [3]_{\mathfrak{R}} = [1]_{\mathfrak{R}} = [-5]_{\mathfrak{R}}$ o $[2]_{\mathfrak{R}} = [11]_{\mathfrak{R}}$, entre una infinidad de posibilidades más.

2.2.3. Aplicaciones entre conjuntos

Veamos ahora otro tipo particular de relaciones entre dos conjuntos con condiciones específicas, conocidas como **aplicaciones** o **funciones**, que se definen de la siguiente manera.

Definición:

Sean A, B conjuntos y f una relación de A en B. Diremos que f es una **aplicación** o **función** si se cumple que $\forall a \in A, \exists! b \in B / afb$, es decir, que para cada uno de los elementos de A existe **un único** elemento de B que está relacionado con A. En este caso, la aplicación suele denotarse por $f:A \rightarrow B$ y se escribe $f(a)=b$ cuando afb .

El conjunto A recibe el nombre de **dominio** de f y se denota por $\text{Dom}(f)$, y el conjunto $f(A)$ definido por $f(A)=\{f(a) / a \in A\}$ recibe el nombre de **codominio, recorrido** o **imagen**, denotado por $\text{Im}(f)$.

Nótese que $\text{Im}(f) \subseteq B$, pero no necesariamente se da la igualdad.

Ejemplos:

Tomamos $A=\{1, 2, 3\}$ y $B=\{1, 3, 5, 8\}$.

1. Sea $f:A \rightarrow B$ de forma que $f(1)=3$, $f(2)=5$ y $f(3)=3$. Entonces f es una aplicación, ya que a cada elemento de A le corresponde un único elemento de B. En este caso, $\text{Dom}(f)=A=\{1, 2, 3\}$ e $\text{Im}(f)=\{3, 5\} \subset B$.

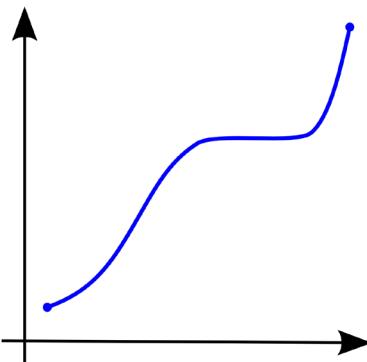


Figura 5. Esquema gráfico de la aplicación.

2. Tomemos $f:A \rightarrow B$ tal que $f(1)=1$, $f(1)=3$, $f(2)=5$ y $f(3)=8$. Entonces f **no** es una aplicación, puesto que a $1 \in A$ no le corresponde un único elemento de B, ya que tiene asignados más de un valor de B, el 1 y el 3.

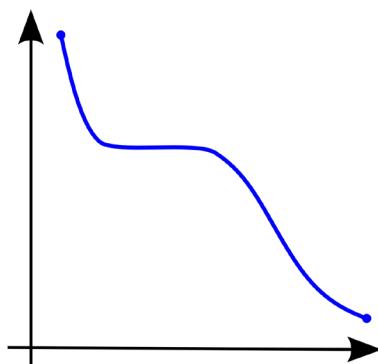


Figura 6. No es una aplicación, ya que el elemento 1 de A se asigna a dos elementos de B.

3. Consideremos $f:A \rightarrow B$ tal que $f(1)=5$ y $f(3)=8$. Entonces f **no** es una aplicación, ya que el valor $2 \in A$ no tiene asignado ningún valor de B a través de f .

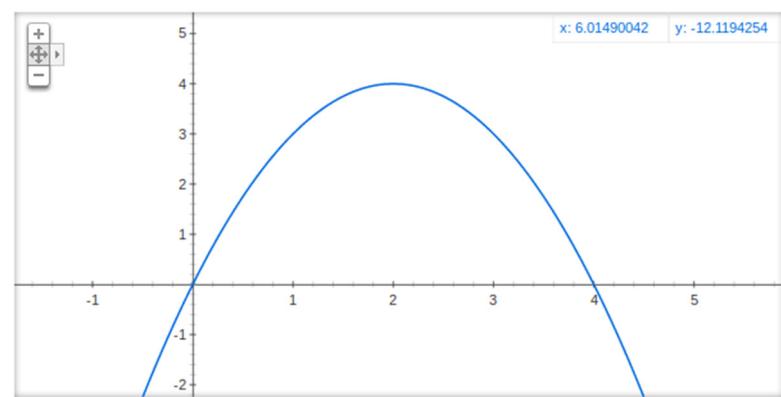


Figura 46. No es una aplicación, ya que el elemento 2 de A no está asignado a ningún elemento de B .

2.2.3.1. Aplicaciones inyectivas, sobreyectivas y biyectivas

Dependiendo de lo que ocurre en el codominio, las aplicaciones pueden clasificarse en tres tipos: inyectivas, sobreyectivas (o suprayectivas) y biyectivas. Veremos con detalle los tres tipos a continuación.

Definición:

Sea $f:A \rightarrow B$ una aplicación. Diremos que f es **inyectiva** si $\forall a, b \in A, f(a)=f(b) \rightarrow a=b$. Dicho de otra forma, una aplicación es inyectiva cuando no hay dos elementos distintos en A a los que f lleve al mismo elemento de B .

Ejemplos:

1. La aplicación del primer ejemplo anterior (véase la figura 5) no es inyectiva, ya que hay dos elementos de A , 1 y 3, a los que f lleva a un mismo elemento de B : el 3.
2. Manteniendo A y B como en los ejemplos anteriores, sea $f:A \rightarrow B$ tal que $f(1)=1, f(2)=3$ y $f(3)$. Entonces f es inyectiva, puesto que no hay ningún valor de f repetido.

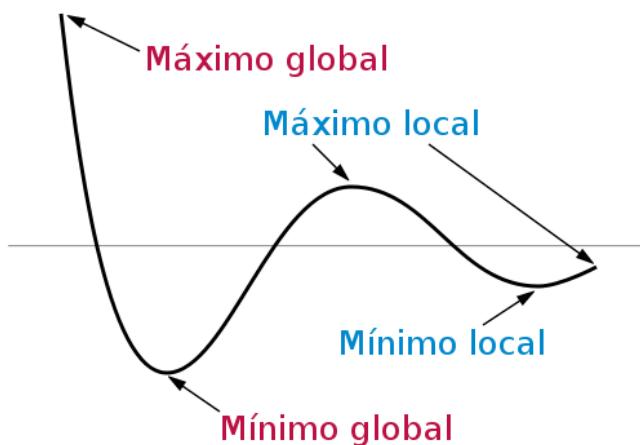


Figura 47. Es inyectiva, ya que todos los elementos de A van a parar a diferentes elementos de B .

Propiedad:

Sea $f:A \rightarrow B$ una aplicación inyectiva, con A, B conjuntos finitos. Entonces $|A|=|f(A)|$. En particular, como $f(A) \subseteq B$, se deduce que $|A| \leq |B|$.

Definición:

Sea $f:A \rightarrow B$ una aplicación. Se dice que f es **sobreyectiva** o **suprayectiva** si $f(A)=B$, o lo que es lo mismo, $\forall b \in B \exists a \in A / f(a)=b$. Es decir, una aplicación es sobreyectiva cuando cada elemento de B tiene asociado al menos un elemento de A al que envía f .

Ejemplos:

1. La aplicación del segundo apartado del ejemplo anterior es inyectiva, pero **no** es sobreyectiva, puesto que f no envía ningún elemento de A al elemento $5 \in B$ (véase la figura 8).
2. Tomemos $A=\{a, \beta, y, \delta\}$ y $B=\{-1, 0, 1\}$. Definimos $f:A \rightarrow B$ de forma que $f(a)=-1$, $f(\beta)=1$, $f(y)=1$ y $f(\delta)=0$. Entonces f es sobreyectiva, ya que cada elemento de B tiene asociado al menos un elemento de A .

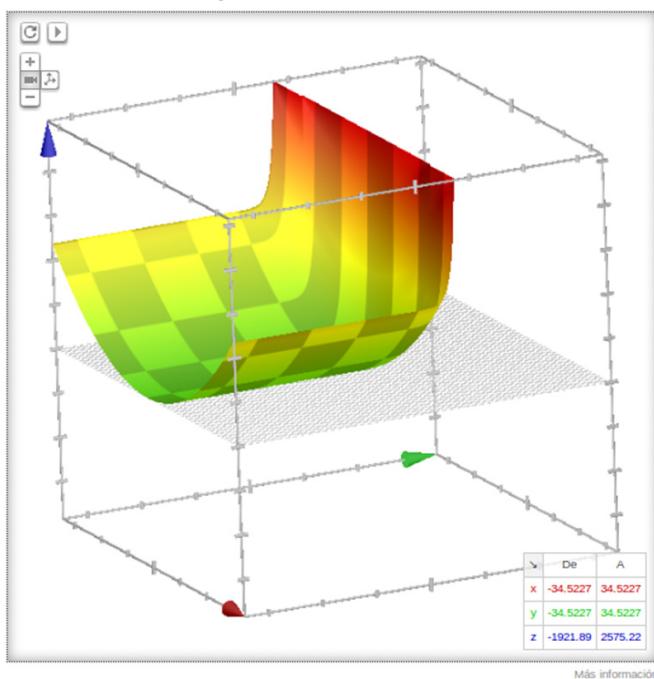


Figura 48. Es sobreyectiva, puesto que cada elemento de B tiene asociado al menos un elemento de A que va a parar a este.

Propiedad:

Sea $f:A \rightarrow B$ una aplicación sobreyectiva, con A, B conjuntos finitos. Entonces $|A| \geq |B|$.

Definición:

Sea $f:A \rightarrow B$ una aplicación. Se dice que f es **biyectiva** si es inyectiva y sobreyectiva simultáneamente. Esta condición se traduce matemáticamente como sigue: $\forall b \in B \exists! a \in A / f(a)=b$.

Nótese que esta definición es muy similar a la de aplicación sobreyectiva, con la salvedad de que esta vez se exige la existencia de un único elemento de A que sea enviado a $b \in B$ por la acción de f, lo cual corresponde a la noción de inyectividad. En esencia, lo que viene a representar la noción de biyección es que hay una correspondencia 1 a 1 entre los elementos de A y los elementos de B, es decir, a cada elemento de A le corresponde un único elemento de B y viceversa.

Ejemplo:

Sean $A=\{\alpha, \beta, \gamma\}$, $B=\{-1, 0, 1\}$ y $f:A \rightarrow B$ dada por $f(\alpha)=0$, $f(\beta)=-1$ y $f(\gamma)=1$. Entonces f es biyectiva, ya que por una parte es inyectiva al enviar elementos diferentes de A a elementos diferentes de B y, por otra parte, cada elemento de B tiene asignado un (único) elemento de A vía f.

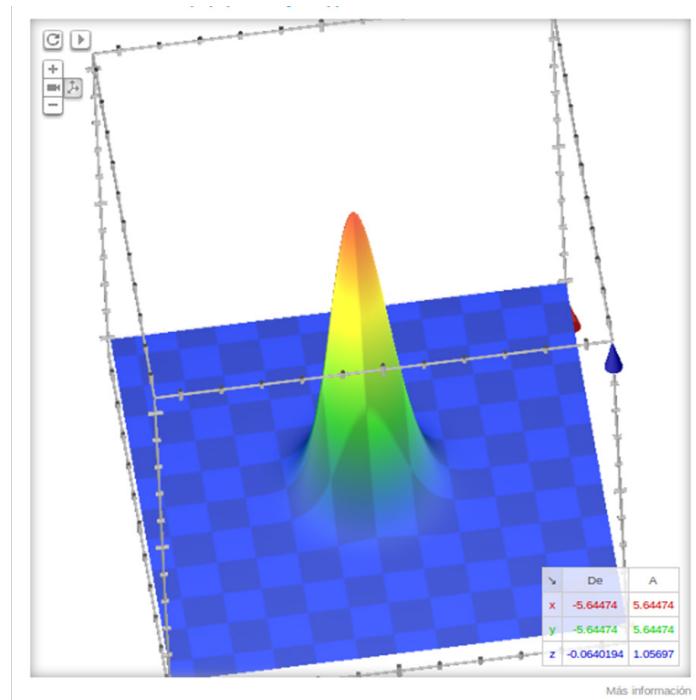


Figura 49. Ejemplo de una aplicación biyectiva.

Propiedad:

Sea $f:A \rightarrow B$ una aplicación biyectiva, con A, B conjuntos finitos. Combinando las propiedades correspondientes para el caso de inyectividad y sobreyectividad, se tiene que entonces $|A|=|B|$.

2.2.3.2. Composición de aplicaciones

Dadas dos aplicaciones, y bajo ciertas condiciones, estas pueden combinarse para formar una nueva aplicación.

Definición:

Sean A, B, C conjuntos cualesquiera, $f:A \rightarrow B$ y $g:B \rightarrow C$. Entonces puede considerarse la **composición** de g con f, $g \circ f:A \rightarrow C$, que viene dada por $a \in A$, $(g \circ f)(a)=g(f(a)) \in C$.

Ejemplo:

Tomemos $A = \{a, \beta, \gamma, \delta\}$, $B = \{-1, 0, 1, 2\}$, $C = \{\rho, \sigma, \tau\}$, $f: A \rightarrow B$ dada por $f(a) = -1$, $f(\beta) = 0$, $f(\gamma) = 0$ y $f(\delta) = 1$, y $g: B \rightarrow C$ dada por $g(-1) = \sigma$, $g(0) = \rho$, $g(1) = \sigma$ y $g(2) = \tau$.

Calculemos $g \circ f: A \rightarrow C$; para ello, tendremos que averiguar su valor en cada uno de los elementos de A :

- $(g \circ f)(a) = g(f(a)) = g(-1) = \sigma$.
- $(g \circ f)(\beta) = g(f(\beta)) = g(0) = \rho$.
- $(g \circ f)(\gamma) = g(f(\gamma)) = g(0) = \rho$.
- $(g \circ f)(\delta) = g(f(\delta)) = g(1) = \sigma$.

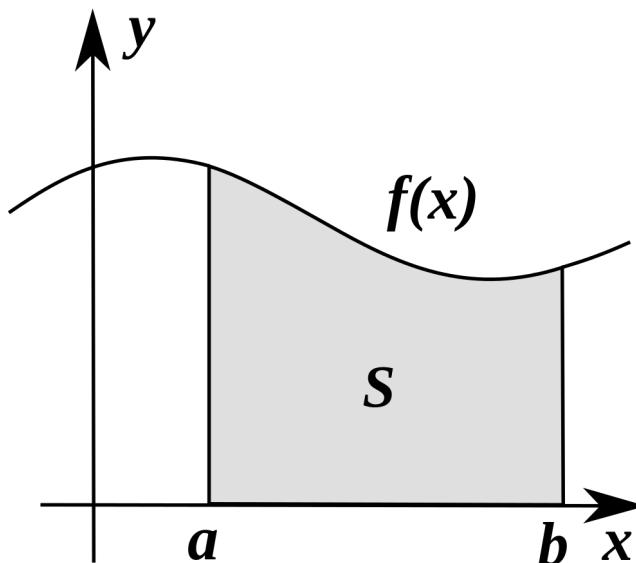


Figura 50. Esquema gráfico acerca de cómo actúa la composición de f con g .

Propiedades:

Sean A, B, C tres conjuntos cualesquiera, y $f: A \rightarrow B$ y $g: B \rightarrow C$. Entonces:

1. Si f y g son inyectivas, entonces $g \circ f$ también lo es.
2. Si f y g son sobreyectivas, entonces $g \circ f$ también lo es.
3. De las dos propiedades anteriores se deduce que si f y g son biyectivas, entonces $g \circ f$ también lo es.

2.2.3.3. Aplicación inversa

Definición:

Sea A un conjunto cualquiera. La aplicación **identidad**, $\text{Id}_A: A \rightarrow A$, es aquella que viene dada por. $\text{Id}_A(a) = a \forall a \in A$.

La noción de aplicación biyectiva junto con el concepto de composición de aplicaciones y el de aplicación identidad dan lugar a otra aplicación asociada cuyo estudio es interesante y cuya existencia viene dada por el siguiente resultado.

Teorema:

Sean A, B dos conjuntos cualesquiera. Entonces $f:A \rightarrow B$ es biyectiva si y solo si $\exists!g:B \rightarrow A$ tal que $g \circ f = f \circ g = \text{Id}_A$.

Definición:

La (única) aplicación g encontrada en el resultado anterior se denomina **aplicación inversa** de f y se denota por $g=f^{-1}$.

La forma de obtener esta aplicación es sencilla, como muestra el ejemplo siguiente.

**Ejemplo**

Sea $A=\{\alpha, \beta, \gamma\}$, $B=\{\rho, \sigma, \tau\}$ y $f:A \rightarrow B$ tal que $f(\alpha)=\tau$, $f(\beta)=\sigma$ y $f(\gamma)=\rho$. Entonces f^{-1} se calcula enviando cada elemento de B al correspondiente elemento de A al que se le asocia f. Por tanto, se tiene $f^{-1}(\rho)=\gamma$, $f^{-1}(\sigma)=\beta$ y $f^{-1}(\tau)=\alpha$.

2.2.3.4. Numerabilidad

Las biyecciones entre conjuntos guardan una estrecha relación con el cardinal de un conjunto. Tanto es así que, en el caso de una biyección entre conjuntos finitos, se concluye que su cardinal ha de ser igual. Así pues, de algún modo cabe pensar que si existe una biyección entre conjuntos, entonces estos son del mismo “tamaño” en cuanto a número de elementos.

Ahora bien, ¿qué ocurre con los conjuntos infinitos? ¿Son todos los conjuntos infinitos igual de “grandes”? Veremos que la respuesta es negativa utilizando biyecciones. Antes de ello, conviene establecer un conjunto infinito de referencia; en este caso tomaremos los números naturales, N.

Definición:

Sea A un conjunto infinito. Diremos que A es un conjunto infinito **numerable** si $\exists F:N \rightarrow A$ biyectiva.

La definición anterior viene a decir, con menos rigor, que un conjunto es numerable cuando todos y cada uno de sus elementos se pueden indexar, sin repetirse y sin dejar ninguno de ellos por listar, mediante índices que varían a lo largo de todos los números naturales.

Ejemplos:

1. Obviamente, N es numerable, puesto que la aplicación $\text{Id}_N:N \rightarrow N$ dada por $\text{Id}_N(n)=n$, $\forall n \in N$ es una biyección.
2. Veamos que el conjunto de números pares, $2N=\{2n / n \in N\}$, es también numerable. En efecto, consideremos la aplicación $f:N \rightarrow 2N$ dada por $f(n)=2n$ y veamos que es biyectiva. Por una parte es inyectiva, ya que si $f(n)=f(n') \rightarrow 2n=2n' \rightarrow n=n'$. Por otra parte, es suprayectiva. Efectivamente, si tomamos $m \in 2N$ entonces m es par y por tanto puede escribirse como $m=2n$, para un cierto $n \in N$. Si consideramos ese mismo valor n, se tiene que $f(n)=2n=m$. Por tanto, f es biyectiva y deducimos con ello que $2N$ es numerable.

3. Comprobemos que Z también es numerable. En efecto, basta con definir $f:N \rightarrow Z$ de la siguiente forma, para cada $n \in N$:

$$f(n) = \begin{cases} \frac{n}{2}, & n \text{ par} \\ -\frac{n-1}{2}, & n \text{ impar} \end{cases}$$

Puede comprobarse sin demasiada dificultad que esta aplicación es biyectiva.

4. El conjunto de los números racionales, Q , también es numerable. La forma en la que se argumenta en este caso es identificando el numerador y denominador de los elementos de Q , vistos como fracciones, con un subconjunto de $Z \times Z$, probando previamente que este último es también numerable.
5. El conjunto de los números reales, R , no es numerable. Cantor fue el primero en probar, hacia 1891, esta afirmación utilizando su argumento de la diagonal, lo que le permitió obtener una demostración muy elegante.

Para más detalles acerca de los últimos ejemplos y sobre cuestiones relativas a la numerabilidad en general, puede consultarse el libro de Bruno Juliá-Díaz y Montserrat Guilleumas *Análisis matemático de una variable*.

2.3. Combinatoria

A continuación se introducirán técnicas útiles para contar, haciendo uso de la teoría de conjuntos. Para ello, se partirá de unos principios básicos relativos a la cardinalidad de conjuntos, para luego dar paso a las cuentas más complicadas.

Se aprovechará también para conectar los conceptos abstractos inherentes a la teoría de conjuntos con situaciones de conteo rutinarias que permitirán establecer una relación y un sentido práctico a su estudio.

2.3.1. Principios de cardinalidad

2.3.1.1. Principio de la adición

Sean A, B dos conjuntos. El principio de cardinalidad dice que si ambos conjuntos son disjuntos, es decir, si no tienen ningún elemento en común, entonces el cardinal de su unión es la suma de sus cardinales. En otras palabras, si $A \cap B = \emptyset$ entonces:

$$|A \cup B| = |A| + |B|.$$

Este principio resulta sencillo de entender, ya que si dos conjuntos no tienen ningún elemento en común, esto quiere decir que al unirlos no está “quitándose” ningún elemento de A ni de B , puesto que no hay ninguno repetido. Por ello, los elementos de $A \cup B$ comprenden **todos** los de A y **todos** los de B , siendo por tanto el número total de elementos los de A más los de B .

Esto puede generalizarse a más de un conjunto de la siguiente forma: si ahora, en lugar de tener dos conjuntos, tenemos una colección finita de n conjuntos A_1, A_2, \dots, A_n , de manera que son disjuntos dos a dos, es decir, $A_i \cap A_j = \emptyset$ si $i \neq j$, entonces se tiene:

$$|A_1 \cup A_2 \cup \dots \cup A_n| = |A_1| + |A_2| + \dots + |A_n|.$$



Ejemplo

De entre todos los alumnos presentados a un examen, han aprobado 34, mientras que 15 de ellos han suspendido. ¿Cuántos alumnos se presentaron en total?

Solución:

Sea A el conjunto formado por los alumnos que han aprobado y B el conjunto de alumnos que han suspendido. Por tanto, se tiene que $|A|=34$ y $|B|=15$. Naturalmente, $A \cup B$, la unión de los estudiantes que han aprobado y los que han suspendido, constituye el conjunto formado por la totalidad de alumnos presentados. Asimismo, $A \cap B = \emptyset$, ya que no se puede aprobar y suspender a la vez, por lo que no puede existir ningún alumno que haya aprobado y suspendido simultáneamente. Por tanto, estamos en condiciones de aplicar el principio de adición:

$$|A \cup B| = |A| + |B| = 34 + 15 = 49.$$

Por tanto, se presentaron un total de 49 alumnos.

2.3.1.2. Principio de Dirichlet

El principio de Dirichlet, también conocido como principio del palomar o principio de las cajas, es un caso particular del principio de adición e indica que si tenemos n conjuntos, cada uno de los cuales contiene como mucho un elemento, entonces el total de elementos, m , cumple $m \leq n$; o lo que es lo mismo, hay menos elementos que cajas.



Ejemplo

El principio de Dirichlet garantiza que, por ejemplo, en una reunión de 8 personas hay al menos una coincidencia en el día de la semana en el que cada una de ellas nació. En efecto, llamemos A_1, A_2, \dots, A_7 a los conjuntos, de forma que en A_i se encuentran las personas que nacieron el día de la semana en la posición i -ésima, es decir, A_i es el conjunto formado por las personas que nacieron un lunes, A_2 el de aquellas que nacieron en martes, y así sucesivamente hasta llegar a A_7 , asociado al domingo. Tenemos en total 7 conjuntos, y si suponemos que no hay ninguna coincidencia, esto quiere decir que los conjuntos anteriores están formados por una única persona o ninguna, por lo que, teniendo en cuenta el principio de Dirichlet, el número de personas sería menor que el de conjuntos, esto es, $8 \leq 7$ lo cual es una contradicción.

Este principio, en apariencia básico y elemental, tiene múltiples aplicaciones en informática. Por ejemplo, sirve para demostrar que no existe ningún algoritmo de hashing capaz de evitar colisiones, en tanto que el número posible de valores que puede tomar un vector excede a menudo el número de sus índices. Asimismo, también es aplicable en el caso de los algoritmos de compresión sin pérdida para demostrar que, si este es capaz de lograr un archivo más pequeño que el original, también existirá otro archivo de entrada que el algoritmo haga irremediablemente más grande, ya que si, de lo contrario, el algoritmo hiciese todos los archivos más pequeños, por el principio de Dirichlet habría dos archivos diferentes cuyas versiones comprimidas serían archivos idénticos, lo cual generaría una inconsistencia en el algoritmo en cuestión.

2.3.1.3. Principio de inclusión-exclusión

Este principio representa una generalización del principio de adición, que proporciona una fórmula para el cardinal de la unión de conjuntos cuando no son necesariamente disjuntos. En el caso de dos conjuntos A y B genéricos, se tiene que:

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

Observemos que en este caso la fórmula también se corresponde con una intuición directa, puesto que, al calcular el número de elementos en un conjunto resultante de unir dos conjuntos no necesariamente disjuntos para obtener $|A \cup B|$, se cuentan una única vez los elementos comunes de los conjuntos que han formado la unión. No obstante, cuando se calcula la suma del cardinal de los dos conjuntos por separado, $|A| + |B|$, se cuentan por duplicado los elementos en común, lo cual se soluciona restando una vez la cantidad de dichos elementos comunes, que no es más que el cardinal de la intersección, $|A \cap B|$, lo cual responde a la fórmula expuesta previamente.

Este principio adquiere una complejidad creciente cuando el número de conjuntos involucrados aumenta. Por ejemplo, en el caso de tres conjuntos A, B, C se tiene:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

En general, se puede comprobar que para n conjuntos A_1, A_2, \dots, A_n se cumple:

$$|A_1 \cup A_2 \cup \dots \cup A_n| = s_1 - s_2 + s_3 - \dots + (-1)^{n-1} s_n,$$

siendo s_i la suma de los cardinales de todos los conjuntos formados por la intersección entre i conjuntos de entre los n conjuntos totales.

Ejemplo

Un grupo de amigos está formado por 15 estudiantes de Matemáticas y 10 de Física, aunque 3 de ellos estudian simultáneamente ambas titulaciones. ¿Cuántos estudiantes de alguna de estas disciplinas hay en total en el grupo de amigos?

Solución:

Sea A el conjunto de alumnos que estudian Matemáticas y B el conjunto de alumnos que estudian Física. Con esta nomenclatura, la traducción de los datos del problema a lenguaje algebraico es la siguiente: $|A|=15$, $|B|=10$, $|A \cap B|=3$, y se pide calcular $|A \cup B|$; por tanto:

$$|A \cup B|=|A|+|B|-|A \cap B|=15+10-3=22.$$

2.3.1.4. Principio de la multiplicación

El principio de la multiplicación hace referencia al cardinal del producto cartesiano de dos conjuntos A, B, denotado por $A \times B$, que viene dado por lo siguiente:

$$|A \times B|=|A| \cdot |B|.$$

Este principio también es intuitivo, ya que si para cada elemento $a \in A$ fijado existen tantos elementos de la forma (a, b) con $b \in B$ variable como elementos hay en B, es decir, $|B|$, entonces, al variar los elementos de A, que en total son $|A|$, tenemos que el número total de pares de la forma (a, b) es $|A| \cdot |B|$.

Su generalización a n conjuntos A_1, A_2, \dots, A_n es natural y viene dada por lo siguiente:

$$|A_1 \times A_2 \times \dots \times A_n|=|A_1| \cdot |A_2| \cdots |A_n|.$$

Ejemplo

¿De cuántas maneras diferentes podemos personalizar un ordenador al que se le reemplaza el disco duro del sistema si debemos decidir entre HDD o SSD y como sistema operativo entre Ubuntu, Debian o Fedora? Si tomamos A como el conjunto formado por los tipos de disco duro, tenemos $A=\{\text{HDD, SSD}\}$, luego $|A|=2$, y B como el conjunto formado por el tipo de sistema operativo, se tiene $B=\{\text{Ubuntu, Debian, Fedora}\}$, siendo por tanto $|B|=3$. Dado que las duplas de valores posibles vienen dadas por el conjunto $A \times B$, por el principio de la multiplicación se tiene que el número de combinaciones posibles es:

$$|A \times B|=|A| \cdot |B|=2 \cdot 3=6.$$

2.3.2. Permutaciones, combinaciones y variaciones

2.3.2.1. Permutaciones sin repetición

Una permutación sin repetición de un conjunto A de n objetos diferentes consiste en cualquier ordenación que se pueda hacer de estos, teniendo en cuenta que el orden de secuenciación importa. Por ejemplo, algunas de las permutaciones asociadas al conjunto $A=\{a, b, c, d\}$ pueden ser secuencias como abcd, bdca o dcba.

Ahora bien, ¿cuántas permutaciones posibles existen? En el ejemplo anterior, observamos que en la primera posición podemos colocar 4 letras. Una vez hecho esto, para cada una de esas cuatro opciones podemos colocar 3 letras en la segunda posición, ya que una de ellas ya se está utilizando en la primera posición. A continuación, para cada una de las posibilidades anteriores habrá 2 letras posibles en la tercera posición, ya que ya hay dos usadas en las posiciones anteriores. Finalmente, para cada una de las posibilidades generadas, queda una única letra que asignar a la cuarta y última posición. Así pues, las posibilidades son $4 \cdot 3 \cdot 2 \cdot 1=4!=24$.

En general, para el caso de un conjunto de n elementos, se tiene que el número de permutaciones posibles para sus elementos se denota por P_n y es:

$$P_n = n!$$

2.3.2.2. Permutaciones con repetición

Dados n elementos de r tipos diferentes, con n_i número de elementos del tipo i , $1 \leq i \leq r$, una permutación con repetición es aquella consistente en una reordenación de estos n elementos, cuya cantidad viene dada por lo siguiente:

$$P_{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! \cdot n_2! \cdots n_r!}$$

Ejemplo

De entre todos los alumnos presentados a un examen, han aprobado 34, mientras que 15 de ellos han suspendido. ¿Cuántos alumnos se presentaron en total?

Solución:

Observemos que en este caso aparecen elementos repetidos, como es el 1, que aparece 2 veces, o el 2, que aparece 3 veces, mientras que el 3 solo aparece una vez. Así pues, tenemos $n=6$ dígitos totales de $r=3$ tipos diferentes (1, 2 y 3), habiendo $n_1=2$ unos, $n_2=3$ doses y $n_3=1$ treses. Por tanto, el número de reordenaciones posibles es:

$$P_6^{2,3,1} = \frac{6!}{2! \cdot 3! \cdot 1!} = 120$$

2.3.2.3. Combinaciones sin repetición

Una combinación sin repetición consiste en tomar k elementos dentro de un conjunto de n elementos diferentes y contar todas las posibles elecciones, teniendo en cuenta que no se pueden repetir elementos, es decir, no se pueden elegir elementos ya escogidos previamente, y donde además el orden no importa. El número de elecciones posibles viene dado por el número combinatorio n sobre k .

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k! \cdot (n - k)!}$$

Ejemplo

¿De cuántas maneras posibles se pueden formar grupos de 3 letras dentro del conjunto de 4 letras $\{a, b, c, d\}$?

Solución:

Tenemos $k=3$ y $n=4$, por lo que el número de elecciones posibles es:

$$C_{4,3} = \binom{4}{3} = \frac{4!}{3! \cdot (4-3)!} = 4,$$

siendo dichas posibilidades abc, abd, acd y bcd.

2.3.2.4. Combinaciones con repetición

Una combinación con repetición se basa en elegir k elementos de un conjunto de n elementos diferentes, donde en esta ocasión puede elegirse un mismo elemento más de una vez y donde el orden sigue sin importar, cuyo número de resultados posibles responde a la fórmula siguiente:

$$CR_{n,k} = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k! \cdot (n-1)!}$$

Ejemplo

Siguiendo con el ejemplo anterior, el número de grupos de 3 dentro del conjunto de 4 letras {a, b, c, d} que se pueden formar si esta vez se permite elegir más de una vez un elemento de este es:

$$CR_{4,3} = \binom{4+3-1}{3} = \binom{6}{3} = 20,$$

siendo todas las posibilidades aaa, aab, aac, aad, abc, abd, acd, bbb, bba, bbc, bbd, bcd, ccc, cca, ccb, ccd, ddd, dda, ddb y ddc.

2.3.2.5. Variaciones sin repetición

Una variación sin repetición consiste en tomar k elementos dentro de un conjunto formado por n elementos diferentes y contar todas las posibles elecciones teniendo en cuenta que no se pueden repetir elementos del conjunto que ya se han elegido previamente, donde, a diferencia de las combinaciones, el orden importa. La expresión para obtener en este caso el número de elementos es:

$$V_{n,k} = n \cdot (n-1) \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

Ejemplo

¿Cuántos números de dos cifras pueden formarse a partir de las cifras de 1234?

Solución:

En este caso tenemos que elegir $k=2$ cifras sobre el conjunto de $n=4$ cifras diferentes, $\{1, 2, 3, 4\}$. Como en esta ocasión el orden de las cifras sí que importa (ya que su alteración genera números distintos) y no se nos permite repetir cifras, tenemos que el número de posibilidades responde a la fórmula de variaciones sin repetición:

$$V_{4,2} = \frac{4!}{2!} = 12$$

siendo los posibles números los siguientes: 12, 13, 14, 21, 23, 24, 31, 32, 34, 41, 42 y 43,

2.3.2.6. Variaciones con repetición

Tomamos de nuevo k elementos en un conjunto formado por n elementos diferentes. Una variación con repetición consiste en considerar todas las posibles elecciones, donde es posible elegir un mismo elemento varias veces y el orden importa. El número de posibilidades responde a la siguiente expresión:

$$VR_{n,k} = n^k.$$



Ejemplo

¿Cuántos números de dos cifras hay que contengan únicamente las cifras 1, 2, 3 y 4?

Solución:

En este caso solo deben contener cifras comprendidas entre 1 y 4, por lo que también se contempla la posibilidad de que se repitan. Así pues, el total de números que pueden formarse es:

$$VR_{4,2} = 4^2 = 16,$$

siendo dichos números: 11, 12, 13, 14, 21, 22, 23, 24, 31, 32, 33, 34, 41, 42, 43 y 44.

2.4. Espacios vectoriales

A continuación estudiaremos las características fundamentales de los espacios vectoriales, para centrarnos más adelante en el caso concreto en el que trabajaremos mayoritariamente, consistente en los hiperplanos reales. Veremos un tipo especial de aplicaciones que pueden formarse entre ellos, las cuales reciben el nombre de aplicaciones lineales, y las matrices que las representan, y finalmente veremos cómo a partir de ellas se construye la noción de diagonalización, un proceso que facilita el cálculo matricial en algunos aspectos.

Definición:

Un espacio vectorial asociado a un cuerpo K , cuyos elementos se denotan como escalares, consiste en un conjunto $V \neq \emptyset$ dotado de una operación interna, suma, $+$, y una operación producto externa \cdot , que involucra elementos de K y de V , de forma que se cumplen las propiedades siguientes:

- 1.** La suma es asociativa:

$$u + (v + w) = (u + v) + w \quad \forall u, v, w \in V.$$

- 2.** La suma es commutativa:

$$u + v = v + u \quad \forall u, v \in V.$$

- 3.** Existe un elemento neutro para la suma:

$$\exists e \in V / u + e = u \quad \forall u \in V.$$

Habitualmente, se suele denotar como $e=0$.

- 4.** Existe un elemento opuesto para la suma:

$$\forall u \in V \exists v \in V / u + v = 0.$$

Habitualmente, suele denominarse como $v=-u$. Asimismo, también suele inducirse la operación “resta” dados $u, v \in V$ cualesquiera, $u-v=u+(-v)$, siendo $-v$ el opuesto de v con respecto a la suma.

- 5.** El producto cumple la propiedad asociativa:

$$a \cdot (b \cdot u) = (a \cdot b) \cdot u, \forall a, b \in K, \forall u \in V.$$

- 6.** Existe un elemento neutro multiplicativo:

$$\exists e \in K / e \cdot u = u \quad \forall u \in V.$$

Habitualmente, suele denominarse como $e=1$.

- 7.** El producto es distributivo respecto a la suma de dos vectores:

$$a \cdot (u + v) = a \cdot u + a \cdot v, \forall a \in K, \forall u, v \in V.$$

- 8.** El producto es distributivo con respecto a la suma de dos escalares:

$$(a + b) \cdot u = a \cdot u + b \cdot u, \forall a, b \in K, \forall u \in V.$$


Ejemplo

- 1.** R^n es un espacio vectorial sobre $R^{m \times n}$, dotado de las siguientes operaciones suma y producto:

Dados $u, v \in R^n$, $u = (u_1, u_2, \dots, u_n)$, $v = (v_1, v_2, \dots, v_n)$, $\lambda \in R$,

$$u + v = (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n),$$

$$\lambda \cdot u = (\lambda u_1, \lambda u_2, \dots, \lambda u_n).$$

- 2.** El grupo de matrices $R^{m \times n}$ también es un espacio vectorial sobre R , mediante sus operaciones suma y producto por escalares habituales.

Dependencia e independencia lineal

Uno de los conceptos fundamentales relacionados con los espacios vectoriales es el de combinación lineal y dependencia lineal, que definimos a continuación.

Definición:

Dados $v_1, a_2 v_2 + \dots + a_k v_k$, una combinación lineal de los k elementos anteriores es cualquier expresión de la forma:

$$a_1 v_1 + a_2 v_2 + \dots + a_k v_k,$$

con $a_1, \dots, a_k \in K$.

Definición:

Sean $v_1, \dots, v_k \in V$. Diremos que los vectores v_1, \dots, v_k son **linealmente independientes** si se cumple que la ecuación:

$$a_1 v_1 + a_2 v_2 + \dots + a_k v_k = 0$$

tiene como única solución $a_1=a_2=\dots=a_k=0$. La afirmación anterior es equivalente a decir que no hay ningún vector de la lista que pueda escribirse como combinación lineal del resto. En caso contrario, se dice que los vectores son **linealmente dependientes**.

Veamos algunos ejemplos para llevar a la práctica estas nociones, en los que apreciaremos la influencia y utilidad inherentes al estudio de sistemas de ecuaciones lineales.

Ejemplos:

Sea $V = \mathbb{R}^3$ y $K = \mathbb{R}$.

1. Los vectores $v_1 = (1, 0, 2)$ y $v_2 = (-2, 0, 4)$ son linealmente dependientes, puesto que puede escribirse:
 $(-2, 0, -4) = (-2) \cdot (1, 0, 2)$,

por lo que se tiene:

$$2v_1 + v_2 = 0,$$

donde en este caso $0 = (0, 0, 0)$.

Es decir, puede escribirse:

$$a_1 v_1 + a_2 v_2 = 0,$$

con $a_1=1$ y $a_2=-2$.

2. Los vectores $v_1 = (1, -1, 2)$ y $v_2 = (0, 1, 1)$ son linealmente independientes. Veámoslo.

Supongamos que tenemos:

$$a_1 v_1 + a_2 v_2 = 0,$$

es decir:

$$a_1 \cdot (1, -1, 2) + a_2 \cdot (0, 1, 1) = 0,$$

donde aplicando las propiedades correspondientes se tiene:

$$(a_1, -a_1 + a_2, 2a_1 + a_2) = (0, 0, 0).$$

Por tanto, igualando componente a componente, se tiene un sistema de tres ecuaciones con dos incógnitas.

$$\begin{cases} a_1 = 0 \\ a_1 + a_2 = 0 \\ 2a_1 + a_2 = 0 \end{cases}$$

Para ver si en efecto son linealmente independientes, tenemos que no existe ninguna solución más allá de la **trivial** ($a_1=a_2=0$). Y, para ello, deberemos probar que el sistema es compatible determinado.

Estudiemos, pues, los correspondientes rangos de las matrices A y A|b, donde en este caso:

$$A = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 2 & 1 \end{pmatrix}, \quad A|b = \left(\begin{array}{ccc|c} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 2 & 1 & 0 \end{array} \right)$$

Se observa que el menor resultante de eliminar la última fila de A tiene por determinante:

$$\det = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 1$$

Por tanto, $\text{ran}(A)=2$. Por otra parte, se observa que la única parte nueva añadida a la matriz A|b es una columna de ceros. En estas condiciones, puede probarse fácilmente que el rango siempre se mantiene, por lo que $\text{ran}(A|b)=\text{ran}(A)=2$. Obsérvese que esto es coherente con el hecho de que un **sistema lineal de ecuaciones homogéneo**, es decir, con todos los términos independientes nulos, siempre tiene al menos una solución, que es la nula, por lo que el sistema siempre será compatible. Es por ello que se omitirá el estudio del rango de A|b por ser siempre el mismo.

Y, en este caso concreto, es compatible determinado, pues $\text{ran}(A)=2$, siendo 2 el número de incógnitas, por lo que la solución es única, y demuestra lo que queríamos.

- 3.** Veamos que los vectores $v_1=(1, 0, -1)$, $v_2=(-1, -2, 0)$, $v_3=(1, 4, -3)$ son linealmente dependientes.

En efecto, planteamos en primer lugar la ecuación:

$$a_1 v_1 + a_2 v_2 + a_3 v_3 = 0$$

y vemos que tiene más soluciones aparte de la trivial. La relación anterior se traduce así: $(a_1+a_2+a_3, -2a_2+4a_3, -a_1-3a_3) = (0, 0, 0)$,

de la cual se infiere un sistema de 3 ecuaciones con 3 incógnitas al llevarla componente a componente:

$$\begin{cases} a_1+a_2+a_3=0 \\ -2a_2+4a_3=0 \\ -a_1-3a_3=0 \end{cases}$$

Tal y como hemos visto en el ejemplo anterior, basta con estudiar el rango de A, donde A es la matriz de coeficientes del sistema:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & 4 \\ -1 & 0 & -3 \end{pmatrix}$$

Como en este caso es cuadrada, empecemos por estudiar su determinante, en representación de su único menor 3×3 , que es la propia matriz, el cual puede comprobarse que es: $\det(A)=0$.

Asimismo, puede comprobarse con facilidad que hay menores 2×2 con determinante no nulo; por ejemplo, el resultante de eliminar la primera fila y la última columna, de lo cual se obtiene:

$$\det = \begin{pmatrix} 0 & -2 \\ -1 & 0 \end{pmatrix} = -2 \neq 0$$

Se concluye, por tanto, que $\text{ran}(A)=2 < n$, siendo $n=3$ el número de ecuaciones. Esto nos permite concluir que existe más de una solución y, por ende, los vectores son linealmente dependientes.

De hecho, podemos ir más allá y encontrar explícitamente algunas combinaciones lineales que dan como resultado el vector nulo, lo cual resuelve el sistema anterior. Veámoslo.

Dado que la eliminación de la primera fila y la última columna ha dado lugar al menor que ha determinado el rango de la matriz, podemos eliminar la primera ecuación y parametrizar la tercera incógnita, $a_3=\lambda$, de lo que resulta el siguiente sistema de ecuaciones:

$$\begin{cases} -2a_2+4\lambda=0 \\ -a_1-3\lambda=0 \end{cases}$$

con lo cual se cumple, observando la segunda y la primera ecuación, respectivamente, que: $a_1=-3\lambda$, $a_2=2\lambda$.

Así pues, el conjunto de las soluciones posibles es:

$$S=\{(-3\lambda, 2\lambda, \lambda) / \lambda \in \mathbb{R}\}.$$

Por tanto, no solo hay una combinación lineal nula posible, sino infinitas. Por ejemplo, para $\lambda=1$, se tiene que $a_1=-3$, $a_2=2$, $a_3=1$ y, efectivamente, puede comprobarse que:

$$-3v_1 + 2v_2 + v_3 = 0.$$

2.4.1. Subespacios vectoriales

Definición:

Sea V un espacio vectorial sobre un cuerpo K . Diremos que un subconjunto $W \subseteq V$ es un **subespacio** vectorial de V , y se denotará por $W \leq V$, si se cumple que:

$$\forall u, w \in W \quad \forall a, b \in V, \quad a \cdot u + b \cdot w \in W.$$

La condición anterior resume el hecho de que W actúa en sí mismo como un espacio vectorial, heredando las operaciones en V , que envían elementos de W a otros elementos de W , tratándose, por tanto, de una estructura cerrada.

Ejemplos:

1. Sea $V=\mathbb{R}^2$ y $K=\mathbb{R}$. Tomamos:

$$W = \{(\lambda, 0) \in \mathbb{R}^2 / \lambda \in \mathbb{R}\}.$$

De la definición se tiene que W es el subconjunto de \mathbb{R}^2 formado por todos los vectores con segunda componente nula. Veamos que, de hecho, es un espacio vectorial. En efecto, si tomamos $u, w \in W$ y $a, b \in \mathbb{R}$, tenemos que, por una parte, u y w pueden escribirse así:

$$u = (\lambda_0, 0), \quad w = (\mu_0, 0),$$

para algún $\lambda_0, \mu_0 \in \mathbb{R}$. Por tanto:

$$au + bw = (a\lambda_0 + b\mu_0, 0) \in W.$$

- 2.** Siguiendo con el mismo espacio vectorial que antes, consideramos ahora:

$$W = \{(\lambda, 1) / \lambda \in \mathbb{R}\}.$$

Entonces, W no es un subespacio vectorial, ya que, por ejemplo, $u=(1, 1) \in W, v=(0, 1) \in W$, pero tomando $a=1$ y $b=-1$, se tiene:

$$u-v = (1, 0) \notin W.$$

El siguiente ejemplo destaca por su importancia.

Teorema:

El conjunto de las soluciones de un sistema de ecuaciones lineal homogéneo es un subespacio vectorial de \mathbb{R}^n , donde n es el número de incógnitas.

Existen formas de crear subespacios vectoriales con facilidad, utilizando elementos del espacio vectorial principal como elementos generadores.

Teorema:

Sea V un espacio vectorial sobre un cuerpo K y sean $v_1, \dots, v_k \in V$. Entonces, el conjunto definido así: $\langle \{v_1, v_2, \dots, v_k\} \rangle = \{\sum_{i=1}^k \lambda_i v_i / \lambda_i \in K\}$

es un subespacio vectorial.

Definición:

El subespacio vectorial anterior recibe el nombre de subespacio generado por $\{v_1, \dots, v_k\}$.

Ejemplos:

Sean $V = \mathbb{R}^3$ y $K = \mathbb{R}$.

- 1.** Tomamos:

$$v = (1, 1, 1).$$

Entonces:

$$\langle v \rangle = \{\lambda v / \lambda \in \mathbb{R}\} = \{(\lambda, \lambda, \lambda) / \lambda \in \mathbb{R}\}.$$

En este caso, hemos obtenido el subespacio vectorial consistente en los vectores del espacio con sus tres componentes iguales.

2. Sean:

$$v_1 = (1, 2, 0), \quad v_2 = (0, 1, -1).$$

Entonces:

$$\langle\{v_1, v_2\}\rangle = \{\lambda_1 v_1 + \lambda_2 v_2 \mid \lambda_1, \lambda_2 \in \mathbb{R}\} = \{(\lambda_1, 2\lambda_1 + \lambda_2, -\lambda_2) \mid \lambda_1, \lambda_2 \in \mathbb{R}\}.$$

2.4.2. Sistemas libres, generadores y bases

Definición:

Sea V un espacio vectorial sobre un cuerpo K . Un subconjunto $S \subseteq V$ se dice que es un **sistema libre** de vectores si todos ellos son linealmente independientes.

Ejemplos:

Tomando $V = \mathbb{R}^3$ y $K = \mathbb{R}$, según hemos visto en los ejemplos anteriores, se tiene:

- 1.** $\{(1, -1, 2), (0, 1, 1)\}$ es un sistema libre.
- 2.** $\{(1, 0, -1), (1, -2, 0), (1, -4, 3)\}$ no es un sistema libre.

Teorema:

Sea V un espacio vectorial sobre un cuerpo K y $S \subseteq V$ finito. Entonces, si S no es un sistema libre $\exists w \in S$ tal que puede escribirse como combinación lineal de otros elementos de S y se tiene que $\langle S \setminus \{w\} \rangle = \langle S \rangle$. El proceso puede iterarse hasta quitar todos los elementos linealmente dependientes y obtener un sistema libre.

Con este procedimiento, se puede demostrar a su vez el siguiente resultado, lo que es de gran utilidad.

Teorema:

Sea $A \in \mathbb{R}^{m \times n}$. Entonces $\text{rank}(A)$ coincide con el número de filas que resultan de quitar todas las filas linealmente dependientes, hasta que los vectores correspondientes a cada fila formen un sistema libre. Es decir, si se han eliminado k filas linealmente dependientes de las demás, de forma que las $m-k$ filas restantes son un sistema libre, entonces $\text{rank}(A)=m-k$. A su vez, este también coincide con el número de columnas resultante de realizar la misma criba que en el caso de las filas. Es decir, si se han eliminado k' columnas linealmente dependientes de las demás, de forma que las $m-k'$ restantes son un sistema libre, entonces $\text{rank}(A)=n-k'$.

Ahora podríamos preguntarnos si es posible generar el espacio vectorial completo de esta forma o, para ser más específicos, si es posible generararlo con un conjunto infinito de vectores y, si ese es el caso, cuál es el número mínimo de vectores con el que se puede conseguir dicho cometido.

Definición:

Sea V un espacio vectorial sobre un cuerpo K . Un subconjunto $S \subseteq V$ se dice que es un **sistema generador** de vectores si $\langle S \rangle = V$.



Ejemplo

Trabajamos con \mathbb{R}^2 sobre \mathbb{R} .

1. El sistema $S=\{(1, 0), (0, 1), (1, 1)\}$ es generador de \mathbb{R}^2 . En efecto, veamos que $\langle S \rangle = \mathbb{R}^2$. Tomamos $v=(v_1, v_2) \in \mathbb{R}^2$ y veamos que puede escribirse como una combinación lineal de elementos de S . En efecto, si tomamos, por ejemplo, los escalares $\lambda_1=v_1, \lambda_2=v_2, \lambda_3=0$, se tiene que:
 $\lambda_1 \cdot (1, 0) + \lambda_2 \cdot (0, 1) + \lambda_3 \cdot (1, 1) = (v_1, v_2)$.
2. El sistema $S=\{(1, 0), (2, 0)\}$ no es un sistema generador de \mathbb{R}^2 , ya que cualquier combinación lineal de ambos elementos deja la segunda coordenada nula, con lo cual es imposible generar todos los elementos de \mathbb{R}^2 .

Definición:

Sea V un espacio vectorial sobre el cuerpo K y $B \subseteq V$. Diremos que B es **base** de V si B es un sistema libre y generador.

Ejemplos:

1. $B=\{(1, 0), (0, 1)\}$ es base de \mathbb{R}^2 .
2. $B=\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ es base de \mathbb{R}^3 .
3. En general, $B=\{e_i\}_{i=1}^n$, con lo cual:

$$e_i(j) = \begin{cases} 1, & i=j \\ 0, & i \neq j \end{cases}$$

es base de \mathbb{R}^n . Esta base recibe el nombre de **base canónica** de \mathbb{R}^n .

Teorema:

Sea V un espacio vectorial sobre un cuerpo K y B una base finita (como conjunto) de V . Entonces, si B' es otra base de V se tiene que B' es también finita y, además, $|B|=|B'|$.

Este resultado indica que, en caso de encontrar una base con un número finito de elementos de un espacio vectorial, todas las demás bases tienen el mismo número de elementos que esta, lo cual permite definir el concepto de dimensión de un espacio vectorial.

Definición:

Sea V un espacio vectorial sobre un cuerpo K . Si B es una base finita de V , con $|B|=n$, diremos que la dimensión de V es n , y lo denotaremos por $\dim(V)=n$.



Ejemplo

De la definición y las consideraciones anteriores se deduce que $\dim(\mathbb{R}^n)=n, \forall n \in \mathbb{N}$.

Nota:

Obsérvese que el concepto de dimensión es también aplicable a subespacios vectoriales de un espacio vectorial de dimensión finita, pues por sí solos también tienen estructura de espacios vectoriales, con una dimensión menor o igual que el original.

2.4.3. Aplicaciones lineales**Definición:**

Sean U, V espacios vectoriales sobre un mismo cuerpo K . Diremos que una aplicación:

$$f: U \rightarrow V$$

es una **aplicación lineal** si $\forall u_1, u_2 \in U$ y $\forall a_1, a_2 \in K$ se cumple:

$$f(a_1 u_1 + a_2 u_2) = a_1 f(u_1) + a_2 f(u_2).$$

Ejemplo

1. Sea $U = \mathbb{R}^3$, $V = \mathbb{R}^2$ y $K = \mathbb{R}$. Entonces la aplicación $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ dada por $f(x, y, z) = (x+y, 2z-x)$ es lineal. En efecto, si tomamos $(x, y, z), (x', y', z') \in \mathbb{R}^3$ y $a, b \in \mathbb{R}$. tenemos, por una parte:
 $f(a \cdot (x, y, z) + b(x', y', z')) = f(ax + bx', ay + by', az + bz') = (ax + bx' + ay + by', 2(az + bz') - (ax + bx')) = (ax + bx' + ay + by', 2az + 2bz' - ax - bx').$

Por otra parte:

$$af(x, y, z) + bf(x', y', z') = a(x+y, 2z-x) + b(x'+y', 2z'-x') = (ax+ay+bx'+by', 2az-ax+2bz'-bx') = (ax+ay+bx'+by', 2az+2bz'-ax-bx').$$

2. Sea $U = \mathbb{R}^2$, $V = \mathbb{R}^2$ y $K = \mathbb{R}$. Consideramos ahora la aplicación $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $f(x, y) = (x, x^2-y)$. Entonces f no es lineal, ya que, por ejemplo, tomando $(2, 0), (0, 0) \in \mathbb{R}^2$ y $2, 0 \in \mathbb{R}$, tenemos, por una parte:

$$f(2 \cdot (2, 0) + 0 \cdot (0, 0)) = f(4, 0) = (4, 16).$$

Por otra parte:

$$2f(2, 0) + 0f(0, 0) = 2 \cdot (2, 4) + 0 \cdot (0, 0) = (4, 8).$$

2.4.3.1. Representación matricial

En adelante, nos centraremos en las aplicaciones lineales sobre los espacios vectoriales reales, es decir, de la forma $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$. El primer resultado importante a este respecto es que hay una conexión íntima entre las matrices y las aplicaciones lineales. Por comodidad con la notación matricial, listaremos las componentes de los vectores por columnas, es decir, dado $v \in \mathbb{R}^n$, escribiremos:

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

De la misma forma, escribiremos en columnas las componentes de la imagen de f .

Teorema:

Toda aplicación lineal $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ tiene una matriz que la representa, que además es única, es decir, $\exists! A \in \mathbb{R}^{m \times n}$, de modo que $\forall v \in \mathbb{R}^n$ puede escribirse:

$$f(v) = A \cdot v.$$

 **Ejemplo**

Sea $f: \mathbb{R}^3 \rightarrow \mathbb{R}^4$ dada por lo siguiente:

$$f(x, y, z) = \begin{pmatrix} x + y \\ y - z \\ 2z + y + z \\ 3z \end{pmatrix}$$

Entonces f viene representada por A , donde:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 2 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix}$$

puesto que, dado $v = (x, y, z) \in \mathbb{R}^3$, tenemos:

$$A \cdot v = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 2 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + y \\ y - z \\ 2z + y + z \\ 3z \end{pmatrix} = f(x, y, z)$$

2.4.3.1.1. Operaciones elementales entre matrices y propiedades

Entendemos como operaciones elementales matriciales para matrices $A_{m \times n}$ a cualquiera de las siguientes:

1. Alternar entre sí dos filas o columnas

Siendo C_i y C_j dos columnas de una matriz $A_{m \times n}$, tenemos que podemos realizar un intercambio entre ellas $C_i \leftrightarrow C_j$. Para el caso de las filas sería análogo al de las columnas.

Siendo A una matriz de la siguiente forma:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,j} \dots a_{1,i} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,j} \dots a_{2,i} \dots a_{2,n} \\ \vdots & \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,j} \dots a_{m,i} \dots a_{m,n} \end{pmatrix}$$

Realizaremos un intercambio en las columnas $C_i = (a_{1,i}, a_{2,i} \dots a_{m,i})$ y $C_j = (a_{1,j}, a_{2,j} \dots a_{m,j})$ tal que A' :

$$A' = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,i} \dots a_{1,j} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,i} \dots a_{2,j} \dots a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,i} \dots a_{m,j} \dots a_{m,n} \end{pmatrix}$$

Ejemplo

Tenemos la siguiente matriz $A_{3 \times 3}$:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix}$$

Si cambiamos las columnas C_2 y C_3 obtenemos la siguiente matriz A' :

$$A' = \begin{pmatrix} 0 & 3 & 1 \\ -1 & 4 & -2 \\ 1 & 2 & 3 \end{pmatrix}$$

2. Multiplicar por un escalar una fila o columna

Siendo r un escalar tal que $r \in \mathbb{R}$ y C_i una columna perteneciente a la matriz $A_{m \times n}$ podemos multiplicar C por dicho escalar tal que $C_i \rightarrow r \cdot C_i$. Para el caso de las filas sería análogo al de las columnas.

Siendo A una matriz de la siguiente forma:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,i} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,i} \dots a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,i} \dots a_{m,n} \end{pmatrix}$$

Realizaremos la operación en la columna $C_i = (a_{1,i}, a_{2,i} \dots a_{m,i})$ tal que $r \cdot C = (r \cdot a_{1,i}, r \cdot a_{2,i} \dots r \cdot a_{m,i})$ de forma que obtengamos una nueva matriz A' :

Ejemplo

$$A' = \begin{pmatrix} a_{1,1} & a_{1,2} \dots r \cdot a_{1,i} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots r \cdot a_{2,i} \dots a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,i} \dots a_{m,n} \end{pmatrix}$$

Tenemos la siguiente matriz $A_{3 \times 3}$:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix}$$

Si aplicamos $r=2$ en la columna C_2 obtenemos la siguiente matriz A' :

$$A' = \begin{pmatrix} 0 & 2 & 3 \\ -1 & -4 & 4 \\ 1 & 6 & 2 \end{pmatrix}$$

3. Sumar a una fila o columna, otra fila o columna multiplicada por un escalar

Siendo r un escalar tal que $r \in \mathbb{R}$ y C_i y C_{kj} dos columnas de una matriz $A_{m \times n}$, tenemos que podemos sumar una columna multiplicada por un escalar a otra tal que $C_i \rightarrow C_i + rC_j$. Para el caso de las filas sería análogo al de las columnas.

Siendo A una matriz de la siguiente forma:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,j} \dots a_{1,i} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,j} \dots a_{2,i} \dots a_{2,n} \\ \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,j} \dots a_{m,i} \dots a_{m,n} \end{pmatrix}$$

Realizaremos la operación en la columna $C_i = (a_{1,i}, a_{2,i} \dots a_{m,i})$ tal que $C_i \rightarrow C_i + rC_j = (a_{1,i} + ra_{1,j}, a_{2,i} + ra_{2,j} \dots a_{m,i} + ra_{m,j})$ de forma que obtengamos una nueva matriz A' :

$$A' = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,j} + ra_{1,i} \dots a_{1,j} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,j} + ra_{2,i} \dots a_{2,j} \dots a_{2,n} \\ \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,j} + ra_{m,i} \dots a_{m,j} \dots a_{m,n} \end{pmatrix}$$

Ejemplo

Tenemos la siguiente matriz $A_{3 \times 3}$:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix}$$

Si aplicamos $r=2$ en la columna C_1 y lo sumamos en C_3 obtenemos la siguiente matriz A' :

$$A' = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 2 \\ 1 & 3 & 4 \end{pmatrix}$$

4. Suma de matrices

Para dos matrices $A_{m \times n}$ y $B_{m \times n}$ (siendo estas de la misma dimensión), la suma de ambas resultará en una nueva matriz con idéntica dimensión $C_{m \times n}$ de forma que para cualquier elemento de esta obtendremos como resultado:

$$c_{i,j} = a_{i,j} + b_{i,j}$$

De forma que para dos matrices A y B tales que:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,n} \\ \vdots & \vdots \quad \ddots \quad \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,n} \end{pmatrix} \text{ y } B = \begin{pmatrix} b_{1,1} & b_{1,2} \dots b_{1,n} \\ b_{2,1} & b_{2,2} \dots b_{2,n} \\ \vdots & \vdots \quad \ddots \quad \vdots \\ b_{m,1} & b_{m,2} \dots b_{m,n} \end{pmatrix}$$

Obtenemos:

$$C = \begin{pmatrix} a_{1,1}+b_{1,1} & a_{1,2}+b_{1,2} & \dots & a_{1,n}+b_{1,n} \\ a_{2,1}+b_{2,1} & a_{2,2}+b_{2,2} & \dots & a_{2,n}+b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}+b_{m,1} & a_{m,2}+b_{m,2} & \dots & a_{m,n}+b_{m,n} \end{pmatrix}$$

Ejemplo

Siendo las matrices A y B:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix} \text{ y } B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 8 & 4 \\ 2 & -1 & 4 \end{pmatrix}$$

Cuando las sumemos obtendremos la matriz C tal que:

$$C = \begin{pmatrix} 1 & 3 & 6 \\ -1 & 6 & 8 \\ 3 & 2 & 6 \end{pmatrix}$$

5. Producto de una matriz por un escalar.

Siendo r un escalar tal que $r \in \mathbb{R}$ y con una matriz $A_{m \times n}$, la multiplicación entre ambos resultará en una nueva matriz para la que cualquier elemento de esta obtendremos como resultado:

$$a'_{ij} = r \cdot a_{ij}$$

De forma que para una matriz A tal que:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,j} & \dots & a_{1,i} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,j} & \dots & a_{2,i} & \dots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,j} & \dots & a_{m,i} & \dots & a_{m,n} \end{pmatrix}$$

Obtenemos:

$$A' = \begin{pmatrix} r \cdot a_{1,1} & r \cdot a_{1,2} & \dots & r \cdot a_{1,n} \\ r \cdot a_{2,1} & r \cdot a_{2,2} & \dots & r \cdot a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r \cdot a_{m,1} & r \cdot a_{m,2} & \dots & r \cdot a_{m,n} \end{pmatrix}$$

Ejemplo

Tenemos la siguiente matriz $A_{3 \times 3}$:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix}$$

Si aplicamos $r=2$ a dicha matriz, obtendremos la siguiente matriz A' :

$$A' = \begin{pmatrix} 0 & 2 & 6 \\ -2 & -4 & 8 \\ 2 & 6 & 4 \end{pmatrix}$$

6. Producto entre matrices

Para dos matrices $A_{m \times n}$ y $B_{n \times p}$ (siendo estas coincidentes en la parte de la dimensión respectiva a n), el producto de ambas resultará en una nueva matriz con idéntica dimensión $A_{m \times p}$ de forma que para cualquier elemento de esta obtendremos como resultado:

$$C_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}$$

Ejemplo

Siendo las matrices A y B:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix} \text{ y } B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 8 & 4 \\ 2 & -1 & 4 \end{pmatrix}$$

Cuando las multipliquemos obtendremos la matriz C tal que:

$$C = \begin{pmatrix} 6 & 5 & 16 \\ 5 & -22 & 5 \\ 5 & 24 & 23 \end{pmatrix}$$

7. Matriz transpuesta

Definimos como la transpuesta de una matriz $A_{m \times n}$ a una matriz $A^t_{n \times m}$ tal que las filas de la matriz A son las columnas de la matriz A^t esto es:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,n} \\ \vdots & \ddots \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,n} \end{pmatrix} \rightarrow A^t = \begin{pmatrix} a_{1,1} & a_{2,1} \dots a_{m,1} \\ a_{1,2} & a_{2,2} \dots a_{m,2} \\ \vdots & \ddots \vdots \\ a_{1,n} & a_{2,n} \dots a_{m,n} \end{pmatrix}$$

Ahora repasaremos las propiedades asociadas a dichas operaciones. Por un lado, tenemos las propiedades asociadas a la suma, siendo A, B y C matrices de tipo $M_{m \times n}$ y los escalares $r, s \in \mathbb{R}$:

1. Propiedad conmutativa en la suma de matrices:

$$A+B=B+A$$

Dado que el resultado de la suma de ambas matrices sería tal que para cada elemento $c_{ij}=a_{ij}+b_{ij}$ y $c'_{ij}=b_{ij}+c_{ij}$, por la propiedad conmutativa natural de la suma de dos elementos reales, tenemos que

2. Propiedad asociativa en la suma de matrices:

$$(A+B)+C=A+(B+C)$$

De manera análoga a la anterior propiedad, para cada elemento $d_{ij}=a_{ij}+b_{ij}+c_{ij}$ vamos a cumplir con la propiedad asociativa natural de la suma de forma que el orden de los sumandos no va a alterar el resultado final.

- 3.** Existencia de un elemento neutro con respecto a la suma, tomando para ello una matriz nula o de ceros tal que la dimensión sea exactamente la misma de la matriz A:

$$A+O=A$$

Para cada elemento $a'_{ij} = a_{ij} + 0 = a_{ij}$. De lo anterior podemos inferir también que existe un elemento opuesto tal que:

$$A+(-A)=0$$

Siendo $-A$ la matriz opuesta de A.

- 4.** Se cumple:

$$r \cdot (A+B) = r \cdot A + r \cdot B$$

Desarrollando la parte izquierda tenemos que para cualquier elemento:

$$c_{ij} = a_{ij} + b_{ij}$$

$$d_{ij} = r \cdot c_{ij}$$

$$d_{ij} = r \cdot (a_{ij} + b_{ij})$$

Desarrollando la parte derecha tenemos que para cualquier elemento:

$$c'_{ij} = r \cdot a_{ij} \text{ y } c''_{ij} = r \cdot b_{ij}$$

$$d'_{ij} = c'_{ij} + c''_{ij} = r \cdot a_{ij} + r \cdot b_{ij} = r \cdot (a_{ij} + b_{ij})$$

De donde obtenemos que $d_{ij} = d'_{ij}$

- 5.** Se cumple:

$$(r+s) \cdot A = r \cdot A + s \cdot A$$

- 6.** Se cumple:

$$r \cdot (s \cdot A) = (r \cdot s) \cdot A$$

Por otro lado tenemos las propiedades asociadas a la suma, siendo A, B y C matrices de tipo $M_{n \times n}$ y los escalares $r \in \mathbb{R}$:

- 1.** Propiedad asociativa para el producto de matrices:

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C$$

Ejemplo con las siguientes matrices A, B, C:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}; B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}; C = \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix} \\ D &= A \cdot (B \cdot C) = A \cdot \left(B \cdot C \right) = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix} \\ D' &= (A \cdot B) \cdot C = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 2 & 9 \end{pmatrix} \cdot \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix} \\ D &= D' \end{aligned}$$

- 2.** Existencia de un elemento neutro con respecto al producto en matrices cuadradas, tomando para ello una matriz unidad o identidad tal:

$$A \cdot I = A$$

Nota:

La matriz unidad es aquella matriz cuadrada cuya diagonal principal está compuesta por unos y por ceros en los elementos tanto por encima como por debajo de la misma:

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

- 3.** Se cumple:

$$r \cdot (A \cdot B) = (r \cdot A) \cdot B = A \cdot (r \cdot B)$$

- 4.** Propiedad distributiva para el producto de matrices:

$$A \cdot (B + C) = A \cdot B + A \cdot C$$

Ejemplo con las siguientes matrices A, B, C:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}; B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}; C = \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix} \\ D &= A \cdot (B + C) = A \cdot \begin{pmatrix} 0 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 7 \\ 3 & 10 \end{pmatrix} \\ D' &= A \cdot B + A \cdot C = \begin{pmatrix} 2 & 6 \\ 2 & 9 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 7 \\ 3 & 10 \end{pmatrix} \\ D &= D' \end{aligned}$$

Por otro lado, tenemos las propiedades asociadas a la transposición de matrices, entre las cuales podemos destacar:

- 1.** La más notable de todas ellas y sobre la que más nos vamos a detener:

$$(A^T)^T = A$$

Siendo la matriz A:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}$$

$$A^T = \begin{pmatrix} a_{1,1} & a_{2,1} & \dots & a_{m,1} \\ a_{1,2} & a_{2,2} & \dots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \dots & a_{m,n} \end{pmatrix}$$

$$(A^t)^t = \begin{pmatrix} a_{1,1} & a_{1,2} \dots a_{1,n} \\ a_{2,1} & a_{2,2} \dots a_{2,n} \\ \vdots & \ddots \vdots \\ a_{m,1} & a_{m,2} \dots a_{m,n} \end{pmatrix} = A$$

Ejemplo:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & \frac{1}{3} & 2 \end{pmatrix}$$

$$A^t = \begin{pmatrix} 0 & -1 & 1 \\ 1 & -2 & 4 \\ 3 & 4 & 2 \end{pmatrix}$$

$$(A^t)^t = \begin{pmatrix} 0 & -1 & 1 \\ -1 & -2 & 4 \\ 1 & \frac{1}{3} & 2 \end{pmatrix} = A$$

2. Se cumplen también:

- a. $(A+B)^t = A^t + B^t$
- b. $(r \cdot A)^t = r \cdot A^t$ para todo $r \in \mathbb{R}$
- c. $(A \cdot B)^t = B^t \cdot A^t$

2.4.3.1.2. Rango, determinante y matriz inversa

Podemos definir el **determinante** de una matriz como un valor representativo de la misma que nos ayudará a averiguar si sus filas o columnas son linealmente independientes entre sí, o para determinar si una matriz tiene inversa. Dicho valor solo es posible de calcular para las matrices cuadradas, es decir, de tipo $M_{n \times n}$.

Para calcular su valor es necesario conocer su dimensión. Vamos a ver los casos más conocidos:

- Matriz de dimensión 1:

$$A = (a)$$

- El valor de su determinante es el valor de su único elemento:

$$|A| = a$$

- Matriz de dimensión 2:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}$$

El valor de su determinante es igual al producto de la diagonal principal menos el producto de la diagonal opuesta:

$$|A| = a_{1,1} \cdot a_{2,2} - a_{1,2} \cdot a_{2,1}$$

- Matriz de dimensión 3:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix}$$

El valor de su determinante se calculará por la regla de Sarrus. Una forma rápida de definirla sería el producto de los elementos de la diagonal principal, más los productos de los elementos de las diagonales paralelas a esta, menos el producto de los elementos de la diagonal opuesta, incluyendo los paralelos. De esta forma tendríamos:

$$|A| = (a_{1,1} \cdot a_{2,2} \cdot a_{3,3}) + (a_{2,1} \cdot a_{3,2} \cdot a_{1,3}) + (a_{3,1} \cdot a_{1,2} \cdot a_{2,3})$$

$$-(a_{3,1} \cdot a_{2,2} \cdot a_{1,3}) - (a_{2,1} \cdot a_{1,2} \cdot a_{3,3}) - (a_{1,1} \cdot a_{3,2} \cdot a_{2,3})$$

Ejemplo:

$$A = \begin{pmatrix} 0 & -1 & 3 \\ -1 & -2 & 4 \\ 1 & \frac{1}{3} & 2 \end{pmatrix}$$

$$|A| = (0 \cdot -2 \cdot 2) + (-1 \cdot 3 \cdot 3) + (1 \cdot 4 \cdot -1) - (1 \cdot -2 \cdot 4) - (1 \cdot 2 \cdot -1) - (0 \cdot 3 \cdot 4)$$

$$|A| = 0 - 0 + 4 + 6 + 2 - 0 = 3$$

- Matriz de dimensión n:

Para los casos en los que el número de dimensiones sea mayor de tres, usaremos la regla de Laplace:

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

Pasemos ahora a estudiar el **rango** de una matriz. Vamos a definir el rango de una matriz como el número de filas o columnas linealmente independientes. Otra forma de definirlo es como la submatriz cuadrada de mayor orden cuyo determinante sea distinto de cero.

De esta forma, con la primera definición podríamos hacer uso del determinante para hallar el rango de una matriz y con la segunda definición, podríamos usar el método de Gauss:

1. Para calcular el rango nos vamos a centrar en las filas. Si tuviéramos el caso de que nuestra matriz tuviera más filas que columnas, podemos calcular su traspuesta y hallar el rango igualmente:

$$\text{rango}(A) = \text{rango}(A^t)$$

2. Haciendo uso de transformaciones elementales, la estrategia a seguir es hacer que por debajo de la diagonal principal, todos sus elementos sean cero.
3. El rango será igual al número de filas no nulas.

Ejemplo 1:

$$A = \begin{pmatrix} 0 & -1 & 3 \\ -1 & -2 & 4 \\ 1 & \frac{1}{3} & 2 \end{pmatrix}$$

a. Por el método del determinante:

Calculamos el determinante de la matriz A:

$$|A| = \begin{vmatrix} 0 & -1 & 3 \\ -1 & -2 & 4 \\ 1 & \frac{1}{3} & 2 \end{vmatrix} = 0 + 1 - 9 + 6 + 2 - 0 = 0$$

Dado que $|A|=0$, tenemos que el rango de la matriz no es 3. Por lo tanto, tenemos que ver si es de rango 2 y para ello nos bastaría con encontrarnos una submatriz A' tal que su determinante fuese distinto de nulo.

Por ejemplo, veamos la submatriz A' :

$$A' = \begin{vmatrix} 0 & -1 \\ -1 & -2 \end{vmatrix} \rightarrow |A'| = \begin{vmatrix} 0 & -1 \\ -1 & -2 \end{vmatrix} = 0 + 1 = 1$$

Como hemos encontrado una submatriz de orden 2 cuyo determinante es distinto de nulo, la matriz A será de rango 2.

b. Por el método de Gauss:

Reordenando las filas tal que $F_1 \leftrightarrow F_3$

$$A = \begin{pmatrix} 1 & 3 & 2 \\ -1 & -2 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

Realizamos la siguiente transformación $F_1+F_2 \leftrightarrow F_2$

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 3 \\ 0 & 1 & 3 \end{pmatrix}$$

Realizamos la siguiente transformación $F_3-F_2 \leftrightarrow F_3$

$$A = \begin{pmatrix} 1 & 3 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

Como podemos ver, tenemos al final de nuestras transformaciones elementales dos filas no nulas y una que sí lo es, luego el rango de nuestra matriz A se confirma que es 2.

Ejemplo 2:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 8 & 4 \\ 2 & -1 & 4 \end{pmatrix}$$

a. Por el método del determinante:

Calculamos el determinante de la matriz A:

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ 0 & 8 & 4 \\ 2 & -1 & 3 \end{vmatrix} = 32 + 16 + 0 - 48 - 4 - 0 = 4$$

Dado que $|A| \neq 0$ tenemos que el rango de la matriz es 3.

b. Por el método de Gauss:

Realizamos la siguiente transformación $F_3-2\cdot F_1 \rightarrow F_3$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 8 & 4 \\ 0 & -5 & -2 \end{pmatrix}$$

Realizamos la siguiente transformación $F_3 + \frac{5}{8}\cdot F_1 \rightarrow F_3$

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 8 & 4 \\ 0 & 0-2+\frac{5}{2} & \end{pmatrix}$$

Como podemos ver, tenemos al final de nuestras transformaciones elementales tres filas no nulas, luego el rango de nuestra matriz A se confirma que es 3.

Por último, vamos a ver qué es la inversa de una matriz y cómo calcularla. Para empezar definiremos la **matriz inversa** A^{-1} de una matriz A como aquella cuyo producto matricial es igual a la matriz identidad I:

$$A \cdot A^{-1} = I$$

Nota:

Existe una relación entre el determinante de una matriz $|A|$ y su inversa A^{-1} y es que si el determinante es nulo, podemos asegurar que esta matriz no es invertible. De ser así, esa matriz pasa a denominarse **matriz singular**.

Para poder realizar el cálculo de la matriz inversa, disponemos de dos métodos:

1. Matriz inversa mediante el cálculo de su determinante y de la matriz adjunta a cada elemento:

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

Donde:

- $|A|$ es el determinante de la matriz A.
- $\text{Adj}(A)$ es la matriz adjunta de A que consiste en otra matriz A' de idéntico orden donde cada elemento de esta nueva matriz se calcula de esta forma:

$$a'_{i,j} = (-1)^{i+j} \cdot |A_{i,j}|$$

- t es la operación de transposición ya vista anteriormente.

Ejemplo:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & 3 & 2 \end{pmatrix}$$

Pasamos a calcular su matriz adjunta:

$$a'_{1,1} = (-1)^{1+1} \cdot |(-2 \cdot 2) - (4 \cdot 3)| = 1 \cdot (-16) = -16$$

$$a'_{1,2} = (-1)^{1+2} \cdot |(-1 \cdot 2) - (1 \cdot 4)| = -1 \cdot (-6) = 6$$

$$a'_{1,3} = (-1)^{1+3} \cdot |(-1 \cdot 3) - (-1 \cdot 2)| = 1 \cdot (-1) = -1$$

$$\text{Adj}(A) = \begin{pmatrix} -16 & 6 & -1 \\ 7 & -3 & 1 \\ 10 & -3 & 1 \end{pmatrix}$$

Calculamos ahora la matriz traspuesta:

$$(\text{Adj}(A))^t = \begin{pmatrix} -16 & 7 & 10 \\ 6 & -3 & -3 \\ -1 & 1 & 1 \end{pmatrix}$$

Y por último, conociendo de ejercicios anteriores que el determinante de A es $|A|=3$ tenemos que la inversa es:

$$A^{-1} = \frac{(\text{Adj}(A))^t}{|A|} = \begin{pmatrix} -\frac{16}{3} & \frac{7}{3} & \frac{10}{3} \\ 2 & -1 & -1 \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

2. Método de Gauss:

Este método consiste en generar una matriz aumentada o extendida $(A | I)$ y a través de operaciones elementales sobre las filas de dicha matriz, iremos convirtiendo la parte correspondiente de la matriz A en una matriz identidad I, de forma que si A es invertible, obtengamos una matriz aumentada de la forma $(I | A^{-1})$.

Ejemplo:

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & -2 & 4 \\ 1 & \frac{1}{3} & 2 \end{pmatrix}$$

Generamos la matriz aumentada:

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 3 & 1 & 0 & 0 \\ -1 & -2 & 4 & 0 & 1 & 0 \\ 1 & \frac{1}{3} & 2 & 0 & 0 & 1 \end{array} \right)$$

Reordenando las filas tal que $F_1 \leftrightarrow F_3$

$$\left(\begin{array}{ccc|ccc} 1 & \frac{1}{3} & 2 & 0 & 0 & 1 \\ -1 & -2 & 4 & 0 & 1 & 0 \\ 0 & 1 & 3 & 1 & 0 & 0 \end{array} \right)$$

Realizamos la siguiente transformación $F_1 + F_2 \rightarrow F_1$

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 6 & 0 & 0 & 1 \\ -1 & -2 & 4 & 0 & 1 & 0 \\ 0 & 1 & 3 & 1 & 0 & 0 \end{array} \right)$$

Realizamos la siguiente transformación $F_2 - F_3 \rightarrow F_2$

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 6 & 0 & 0 & 1 \\ 0 & 0 & 3 & -1 & 1 & 1 \\ 0 & 1 & 3 & 1 & 0 & 0 \end{array} \right)$$

Realizamos la siguiente transformación $F_1 - 3F_2 \rightarrow F_1$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 6 & 0 & 1 & 1 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right)$$

Realizamos la siguiente transformación $F_1 + \frac{16}{3}F_3 \rightarrow F_1$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{16}{3} & \frac{7}{3} & \frac{10}{3} \\ 0 & 1 & 6 & 0 & 1 & 1 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right)$$

Realizamos la siguiente transformación $F_2 - 2 \cdot F_3 \rightarrow F_2$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{16}{3} & \frac{7}{3} & \frac{10}{3} \\ 0 & 1 & 0 & 2 & -1 & -1 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array} \right)$$

Realizamos la siguiente transformación $\frac{1}{3} \cdot F_3 \rightarrow F_3$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{16}{3} & \frac{7}{3} & \frac{10}{3} \\ 0 & 1 & 0 & 2 & -1 & -1 \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right)$$

Con lo que tenemos una matriz aumentada de la forma $(I | A^{-1})$ tal que:

$$A^{-1} = \left(\begin{array}{ccc} -\frac{16}{3} & \frac{7}{3} & \frac{10}{3} \\ 2 & -1 & -1 \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right)$$

2.4.3.2. Autovalores y autovectores

El siguiente resultado nos presenta los conceptos clave para proceder a la diagonalización de matrices.

Definición:

Sea $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ una aplicación lineal. Un **vector propio** o **autovector** $v \neq 0$ asociado a un valor propio o autovalor λ de f es aquel que cumple $f(v) = \lambda v$. Es decir, aquel tal que f lo envía a un múltiplo de sí mismo.

Si A es la matriz que representa a f , podemos escribir la relación de la siguiente forma:

$$A \cdot v = \lambda v.$$

Para obtener un procedimiento que permita plantear las ecuaciones para obtener las componentes de los candidatos a vectores propios y los valores propios en sí, basta con observar que λv puede escribirse como $\lambda I_n \cdot v$, donde I_n es la matriz identidad $n \times n$. Así pues, se tiene:

$$A \cdot v = \lambda I_n \cdot v \rightarrow A \cdot v - \lambda I_n \cdot v = 0 \rightarrow (A - \lambda I_n) \cdot v = 0.$$

Por tanto, si se denota:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{pmatrix}, I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

se tiene que la relación anterior se traduce así:

$$\begin{pmatrix} a_{1,1} - \lambda & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} - \lambda & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} - \lambda \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

Así pues, se trata de un problema de encontrar soluciones no triviales (puesto que, por definición de vector propio, este no puede ser el vector nulo) a un sistema homogéneo. Obsérvese que, en caso de que el sistema fuese determinado, tendría una única solución, y esta obviamente sería la trivial, que no interesa. Por tanto, para encontrar soluciones no triviales **el sistema debe ser compatible indeterminado**.

Así pues, la idea detrás de la búsqueda de valores y vectores propios estriba en el hecho de hallar valores de λ para los cuales la matriz $A - \lambda I_n$ es singular. No obstante, observemos que en este caso la matriz es cuadrada, por lo que la afirmación anterior es equivalente a decir que $\det(A - \lambda I_n) = 0$. Los valores λ que cumplan esa condición serán los valores propios y los vectores resultantes de la solución de dicho sistema compatible indeterminado los vectores propios.

Definición:

La expresión $\det(A - \lambda I_n)$ es un polinomio de grado n que recibe el nombre de **polinomio característico**.

El siguiente resultado es crucial en términos de búsqueda de vectores propios y, en última instancia, para la diagonalización.

Teorema:

El conjunto de soluciones de un sistema homogéneo $n \times n$ es siempre un subespacio vectorial de \mathbb{R}^n . En particular, el conjunto de vectores propios asociados a un valor propio λ de una aplicación lineal $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ es también un subespacio vectorial de \mathbb{R}^n .

Además, si A es la matriz de coeficientes del sistema y S el conjunto de todas las soluciones del sistema homogéneo asociado, que por lo anterior es un subespacio vectorial de \mathbb{R}^n , se cumple que $\dim(S) = n - k$, donde $k = \text{rank}(A)$.

Este resultado, junto con las consideraciones anteriores, motiva la siguiente definición, que es determinante para caracterizar cuándo una matriz es diagonalizable y cuándo no.

Definición:

Sea λ un valor propio asociado a una aplicación lineal $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Entonces se define:

- Dimensión algebraica es la multiplicidad de λ como raíz de la ecuación polinómica asociada al polinomio característico $\det(A - \lambda I_n) = 0$.
- Dimensión geométrica es la dimensión del subespacio vectorial formado por los vectores propios asociados al valor propio λ , es decir, $\dim(V_\lambda)$, donde:
 $V_\lambda = \{v \in \mathbb{R}^n / A \cdot v = \lambda v\}$.

Nota:

Obsérvese que cuando se habla de subespacio vectorial asociado a los vectores propios también se considera el vector nulo, ya que de lo contrario no sería un subespacio.

2.5. Factorización de matrices

En este apartado veremos algunas factorizaciones conocidas para algunos tipos de matrices. Estas factorizaciones pueden ser muy útiles para resolver determinados problemas o para simplificar algunas operaciones entre matrices.

2.5.1. Diagonalización

Definición:

Sea $A \in \mathbb{R}^{n \times n}$. Diremos que A es diagonalizable si $\exists P \in \mathbb{R}^{n \times n}$ regular tal que:

$$P^{-1}AP=D,$$

siendo D una matriz diagonal, es decir, de la forma:

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

El siguiente resultado proporciona una caracterización completa de las matrices diagonalizables, además de describir cómo son las matrices P y D, relacionándolas con los vectores propios y valores propios asociados a A, respectivamente.

Teorema:

Sea $A \in \mathbb{R}^{n \times n}$. Entonces A es diagonalizable solo si tanto la suma de las dimensiones algebraicas como la de las dimensiones geométricas asociadas a todos los valores propios es n y, además, la dimensión algebraica y la dimensión geométrica coinciden.

Además, en ese caso, la matriz D está formada por todos los valores propios (cada uno de ellos repetido tantas veces como su multiplicidad) y la matriz P por vectores propios linealmente independientes colocados por columnas, de forma que el vector propio de la columna i de P está asociado al valor propio de la entrada (i, i) de la matriz D.

Nota:

Del teorema anterior se deduce que, en el caso de que haya exactamente n valores propios diferentes, esto quiere decir que su multiplicidad es 1, y con ello puede comprobarse que la dimensión algebraica asociada coincide con la geométrica, que también es 1, por lo que la matriz resulta diagonalizable.

Ejemplos:

Veamos una serie de ejemplos para ilustrar la forma de proceder a la hora de diagonalizar una matriz y todos los casos que pueden presentarse.

1. Sea:

$$A = \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix}.$$

Vamos a diagonalizar A. El primer paso es calcular las raíces del polinomio característico:

$$\det(A - \lambda I_2) = 0.$$

Se tiene, por tanto:

$$A - \lambda I_2 = \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 5 - \lambda & -3 \\ 6 & -4 - \lambda \end{pmatrix}$$

por lo que:

$$\det(A - \lambda I_2) = (5 - \lambda)(-4 - \lambda) - (-3) \cdot 6 = \lambda^2 - \lambda - 2.$$

Así pues, como $\det(A - \lambda I_2)$, se tiene que:

$$\lambda^2 - \lambda - 2 = 0$$

Por tanto, se obtiene:

$$\lambda = \frac{-(-1) \pm \sqrt{1 + 2 \cdot 1 \cdot (-4)}}{2 \cdot 1} = \frac{1 \pm 3}{2}$$

con lo cual obtenemos dos **raíces simples**:

$$\lambda_1 = -1, \quad \lambda_2 = 2$$

Así pues, se han obtenido 2 raíces, que resultan ser los valores propios asociados a la matriz A. Como su multiplicidad algebraica es 1 y hay 2 en total, que coinciden con el número de filas y columnas de A, se tiene que A es diagonalizable, y podemos escribir la matriz D del siguiente modo:

$$D = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Ahora, para calcular la matriz P, estudiemos los vectores propios asociados a los valores propios obtenidos.

Empecemos por el caso $\lambda_1 = -1$. En ese caso, tenemos que la matriz:

$$A - (-1) \cdot I_2 = A + I_2 = \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 6 & -3 \\ 6 & -3 \end{pmatrix}$$

es singular, pues, como hemos visto, en ese caso su determinante es nulo. Por ese motivo, se tiene que el sistema homogéneo:

$$\begin{pmatrix} 6 & -3 \\ 6 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

o lo que es lo mismo:

$$\begin{cases} 6x - 3y = 0 \\ 6x - 3y = 0 \end{cases}$$

es compatible indeterminado, por lo que existen más soluciones además de la trivial, (0, 0).

Como, por ejemplo, el menor 1×1 asociado a la entrada de la segunda fila y la segunda columna es no nulo, tenemos que $\text{rank}(A)=1$ y podemos eliminar la primera ecuación (que, de hecho, en este caso es la misma que la segunda) y parametrizar la primera incógnita, $x=a$. Así pues, tenemos el siguiente sistema reducido, equivalente al anterior:

$$\{-3y = -6a,$$

por lo que:

$$y = \frac{-6a}{-3} = 2a$$

Así pues:

$$V_{-1} = \{(a, 2a) / a \in \mathbb{R}\}.$$

Por ejemplo, tomando $a=1$, tenemos que un vector propio no nulo es $(1, 2)$ asociado al valor propio $\lambda=-1$.

Continuemos ahora con el caso λ_2 . En este caso, se tiene que la matriz:

$$A - 2I_2 = \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -3 \\ 6 & -6 \end{pmatrix}$$

también es singular, pues su determinante es nulo, tal y como se ha visto. Por tanto, el sistema homogéneo:

$$\begin{pmatrix} 3 & -3 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

o, equivalentemente:

$$\begin{cases} 3x - 3y = 0 \\ 6x - 6y = 0 \end{cases}$$

es compatible indeterminado (se ve con claridad al observar que la segunda ecuación se obtiene a partir de la primera, multiplicándola por 2).

Como todas las entradas de la matriz son no nulas, podemos eliminar la ecuación que queramos, por ejemplo, la segunda, y parametrizar también la variable que queramos, por ejemplo, la segunda, $y=a$, con lo cual obtenemos el sistema reducido equivalente:

$$\{ 3x = 3a,$$

que a su vez permite obtener:

$$y = \frac{3a}{3} = a$$

Así pues:

$$V_2 = \{(a, a) / a \in \mathbb{R}\}.$$

Por tanto, un ejemplo de vector propio no nulo se obtiene tomando $a=1$, por lo que $(1, 1)$ es un vector propio no nulo asociado al valor propio $\lambda=2$.

Así, podemos construir la matriz P poniendo los vectores propios por columnas, siguiendo el orden en el que se han puesto los valores propios en D. Así pues, como en el primer elemento de la diagonal de la matriz D aparece el valor propio -1 y en el segundo aparece el valor propio 2, colocaremos en la primera columna de P el vector propio $(1, 2)$ (asociado al valor propio $\lambda=-1$) y en la segunda columna el vector propio $(1, 1)$ (asociado al valor propio $\lambda=2$):

$$P = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix}.$$

Podemos calcular ahora la inversa de P, que puede comprobarse que es:

$$P^{-1} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix}.$$

Por tanto, hemos encontrado dos matrices D, P, con D diagonal y P invertible, de forma que:

$$P^{-1}AP = D.$$

Veamos que, en efecto, dicha relación se da:

$$(P^{-1}A)P = \left[\begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} 5 & -3 \\ 6 & -4 \end{pmatrix} \right] \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix} = D,$$

por lo que hemos comprobado que el problema está correctamente resuelto.

2. Sea:

$$A = \begin{pmatrix} -2 & -1 & 1 \\ 0 & 1 & 0 \\ -6 & -2 & 3 \end{pmatrix}$$

Veamos si A es diagonalizable y, si es el caso, procedamos a la obtención de D y P. Consideraremos el polinomio característico:

$$A - \lambda I_3 = \begin{pmatrix} -2 - \lambda & -1 & 1 \\ 0 & 1 - \lambda & 0 \\ -6 & -2 & 3 - \lambda \end{pmatrix}$$

Así pues, calculando el determinante y desarrollándolo por la primera columna, se tiene:

$$\begin{aligned} \det(A - \lambda I_3) &= (-2 - \lambda) \det \begin{pmatrix} 1 - \lambda & 0 \\ -2 & 3 - \lambda \end{pmatrix} - 6 \det \begin{pmatrix} -1 & 1 \\ 1 - \lambda & 0 \end{pmatrix} \\ &= (-2 - \lambda)(1 - \lambda)(3 - \lambda) + 6(1 - \lambda) \\ &= (1 - \lambda)[(-2 - \lambda)(3 - \lambda) + 6] = (1 - \lambda)(\lambda^2 - \lambda) = \lambda(1 - \lambda)^2 \end{aligned}$$

Por tanto, la ecuación que hay que resolver es:

$$\lambda(1 - \lambda)^2 = 0,$$

de manera que se obtienen los valores propios $\lambda_1=0$, $\lambda_2=1$, $\lambda_3=1$, puesto que 1 es una raíz doble de la ecuación anterior. Por tanto, la matriz D puede escribirse del siguiente modo:

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

Como hay valores propios que se repiten, todavía no sabemos si A es diagonalizable o no; todo dependerá de si $(V_1)=2$ o no, es decir, de si la multiplicidad geométrica, asociada a la dimensión del subespacio de vectores propios asociado al valor propio $\lambda=1$, coincide con la dimensión algebraica, asociada a la multiplicidad del valor propio $\lambda=1$ como raíz del polinomio característico.

Por una parte, para el autovalor $\lambda=0$ se tiene que:

$$A - 0 \cdot I_3 = A = \begin{pmatrix} -2 & -1 & 1 \\ 0 & 1 & 0 \\ -6 & -2 & 3 \end{pmatrix}$$

es singular, por lo que el sistema de ecuaciones homogéneo asociado:

$$\left\{ \begin{array}{l} -2x - y + z = 0 \\ y = 0 \\ -6x - 2y + 3z = 0 \end{array} \right.$$

es compatible indeterminado. Se tiene que $\text{rank}(A)=2$, pues, por ejemplo, tomando el menor resultante de eliminar la última fila y la última columna se tiene que:

$$\det \begin{pmatrix} -2 & -1 \\ 0 & 1 \end{pmatrix} = -2 \neq 0.$$

por lo que su solución tiene $3-2=1$ parámetro y, por lo anterior, puede eliminarse la última ecuación y parametrizar la última incógnita, $z=a$, de manera que se obtiene el sistema reducido equivalente:

$$\left\{ \begin{array}{l} -2x - y = -a \\ y = 0 \end{array} \right.$$

Sustituyendo $y=0$ en la primera ecuación, se tiene:

$$-2x = -a$$

luego:

$$x = \frac{-a}{-2} = \frac{a}{2}$$

Por tanto, el conjunto de las soluciones es el subespacio formado por los vectores propios asociados al valor propio $\lambda=0$, es decir:

$$V_0 = \left\{ \left(\frac{a}{2}, 0, a \right) / a \in \mathbb{R} \right\}.$$

Así pues, tomando por ejemplo $\alpha=2$ tenemos que $(1, 0, 2) \in V_2$ es un vector propio asociado al valor propio $\lambda=0$.

Por otra parte, para el autovalor $\lambda=1$ tenemos:

$$A - I_3 = \begin{pmatrix} -3 & -1 & 1 \\ 0 & 0 & 0 \\ -6 & -2 & 2 \end{pmatrix}$$

de lo cual se induce un sistema de ecuaciones homogéneo con 2 ecuaciones y 3 incógnitas, pues la segunda fila de A es nula y no aporta ninguna información relevante ($0=0$):

$$\begin{cases} -3x -y +z = 0 \\ -6x -2y +2z = 0 \end{cases}$$

Asimismo, se observa que la segunda ecuación se obtiene fácilmente de la primera multiplicándola por 2, por lo que también es redundante y puede eliminarse, de lo que resulta:

$$\{ -3x -y +z = 0.$$

Así pues, hemos visto que las filas segunda y tercera de A son una combinación lineal de la primera, pues la segunda es igual a la primera multiplicada por 0 y la tercera es igual a la primera multiplicada por 2. Por tanto, como se requieren 2 eliminaciones de filas para que el conjunto de vectores fila sea linealmente independiente, se tiene que $\text{rank}(A)=1$. Se ha utilizado este análisis alternativo porque probar que todos los menores 2×2 tienen determinante nulo es más costoso, aunque sirve igualmente para probar que el rango de A en este caso es 1.

Por otro resultado previo, como V_1 es el conjunto de soluciones del sistema anterior, asociado al valor propio $\lambda=1$, se tiene que $\dim(V_1)=n-\text{rank}(A)=3-1=2$. Así pues, la multiplicidad geométrica coincide con la multiplicidad algebraica asociadas al valor propio $\lambda=1$, siendo 2 en ambos casos, y como ya no queda ningún valor propio más para analizar, se concluye que A es diagonalizable.

Ahora, para obtener V_1 , sabiendo que $\dim(V_1)=2$, parametrizamos dos incógnitas, haciendo, por ejemplo, $x=\alpha$ e $y=\beta$, con lo cual:

$$z = 3\alpha + \beta.$$

Así pues:

$$V_1 = \{(\alpha, \beta, 3\alpha + \beta) / \alpha, \beta \in \mathbb{R}\}.$$

En este caso, debemos tomar 2 vectores propios de V_1 , pues el valor propio asociado tiene multiplicidad 2, con la precaución de que sean linealmente independientes.

Por ejemplo, tomando $\alpha=1$ y $\beta=0$, un vector propio es $(1, 0, 3)$. Tomando ahora $\alpha=0$ y $\beta=1$, otro vector propio es $(0, 1, 1)$, donde puede comprobarse que son linealmente independientes. Así pues, podemos construir P del siguiente modo:

$$P = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 2 & 3 & 1 \end{pmatrix}.$$

así como calcular su matriz inversa, que puede comprobarse que es:

$$P^{-1} = \begin{pmatrix} 3 & 1 & 1 \\ -2 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Así pues, el problema de diagonalización está resuelto y puede comprobarse que, efectivamente:

$$P^{-1}AP = D.$$

3. Consideremos:

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

Veamos que A no es diagonalizable. En efecto, tenemos:

$$A - \lambda I_2 = \begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix},$$

por lo que su polinomio característico es:

$$\det(A - \lambda I_2) = (-\lambda) \cdot (-\lambda) - 1 \cdot (-1) = \lambda^2 + 1.$$

Sin embargo, la ecuación:

$$\det(A - \lambda I_2) = 0$$

no produce ningún valor propio, pues la ecuación polinómica:

$$\lambda^2 + 1 = 0$$

no tiene raíces reales.

Por tanto, se trata de un caso en el que, al necesitarse dos raíces reales para obtener los correspondientes valores propios, no se ha obtenido, de hecho, ninguna.

4. Tomemos ahora:

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

Veamos que esta matriz resulta ser nuevamente no diagonalizable. En efecto, se tiene:

$$A - \lambda I_2 = \begin{pmatrix} -\lambda & 1 \\ 0 & -\lambda \end{pmatrix}.$$

por lo que, si tomamos su polinomio característico, tenemos:

$$\det(A - \lambda I_2) = \lambda^2$$

Así pues, la ecuación asociada, $\det(A - \lambda I_2) = 0$, es equivalente a lo siguiente:

$$\lambda^2 = 0,$$

que tiene por solución 0 como raíz doble, por lo que se obtienen los valores propios $\lambda_1=0$ y $\lambda_2=0$. Por tanto, la dimensión algebraica asociada a este (único) valor propio es 2. Veamos que, no obstante, la dimensión geométrica asociada a dicho valor propio es 1, es decir, $\dim(V_0)=1$.

En efecto, el sistema homogéneo asociado a la matriz de coeficientes:

$$A - 0 \cdot I_2 = A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

es:

$$\begin{cases} y = 0 \\ 0 = 0 \end{cases}$$

Ahora bien, como $\text{rank}(A)=1$ (pues $\det(A)=0$, pero tiene una entrada no nula), se tiene que $\dim(V_k)=1$. De hecho, puede apreciarse en el sistema que $y=0$, por lo que la incógnita y está únicamente determinada y, por consiguiente, solo se admite un parámetro, en este caso para la incógnita x .

Por tanto, A no es diagonalizable.

2.5.2. Factorización LU

Esta factorización se puede aplicar a todas las matrices cuadradas no singulares y permite escribir la matriz como producto de dos matrices (triangular inferior y triangular superior) de forma única. En las fórmulas, $A \in \mathbb{R}^n$ es una matriz cuadrada no singular, es decir, $\det(A) \neq 0$. Así pues, existen dos únicas matrices $L, U \in \mathbb{R}^{n \times n}$, L triangular inferior y U triangular superior, tales que $A=LU$.

Si se conoce la factorización LU de una matriz, entonces es muy sencillo calcular su determinante, dado que el determinante de un producto de dos matrices es el producto de los determinantes de los factores. Además, al ser los factores matrices triangulares, sus determinantes son simplemente el producto de los elementos de las diagonales.

2.5.3. Factorización de Cholesky

La factorización de Cholesky es la descomposición de una matriz simétrica y definida positiva como producto de una matriz triangular inferior y su traspuesta:

$$S=LL^t \text{ para cada } S \in \mathbb{R}^{n \times n} \text{ simétrica y definida positiva.}$$

Esta factorización es única, es decir, la matriz L es única y es un caso particular de la factorización LU para matrices simétricas definidas positivas.

2.5.4. Factorización QR

La factorización QR descompone cada matriz cuadrada A en el producto de una matriz ortogonal Q y una matriz triangular superior R :

$$A=QR.$$

La forma más sencilla de calcular la factorización QR es utilizar el método de Gram-Schmidt para ortogonalizar las columnas de la matriz A. El método de Gram-Schmidt es un algoritmo iterativo que transforma un conjunto de vectores en un conjunto de vectores ortogonales entre sí. A cada paso, el algoritmo modifica uno de los vectores restándole las proyecciones de los otros vectores ya modificados (calculando entonces los complementos ortogonales).

2.5.5. Descomposición en valores singulares (SVD)

Esta descomposición se puede aplicar a todas las matrices $A \in \mathbb{R}^{n \times n}$, también a las no cuadradas. Existen dos matrices ortogonales $U \in \mathbb{R}^{m \times m}$ y $V \in \mathbb{R}^{n \times n}$ tales que:

$$A = U\Sigma V,$$

donde $\Sigma \in \mathbb{R}^{m \times n}$ es una matriz de ceros con solo algunos valores distintos de 0 en la diagonal principal. Estos valores son los valores singulares de A y son iguales a la raíz cuadrada de los autovalores de la matriz $A^t A \in \mathbb{R}^{n \times n}$, que, al ser una matriz simétrica, es siempre diagonalizable.

2.6. Cálculo tensorial

En este apartado vamos a iniciarnos en conceptos básicos sobre el cálculo tensorial, pudiendo definir este como una generalización del álgebra lineal. Un tensor es una interpretación matemática de un concepto puramente físico, ya que son independientes del sistema de referencia. Centrándonos en el concepto matemático, podemos decir que un tensor es una aplicación sobre un espacio vectorial.

Sin embargo, cuando llevamos los tensores a la práctica en el mundo físico y los asociamos a un sistema de referencia, aparece el concepto de “componente del tensor”, que, al contrario de su definición matemática, al estar relacionado con dicho sistema, variará y se verá afectado por las reglas sujetas a este.

En 1916, Albert Einstein introdujo el concepto de notación indicial o indexada en el Convenio de Suma de Einstein. Dicho convenio nos sirve para representar sumatorios de elementos tal que:

$$\sum_{i=1}^5 i^3 = 1^3 + 2^3 + 3^3 + 4^3 + 5^3$$

Según Einstein, podemos prescindir del símbolo de sumatorio dejando tan solo los índices que serían representativos del sumatorio, añadiendo, si hiciera falta, el número de veces que se repite este. En el Convenio, Einstein especificó que esto solo podía ser llevado a cabo para aquellos índices que se repitieran dentro de la expresión que simboliza el sumatorio, siendo denominados estos **“índices mudos”**. Por el contrario, a aquellos que aparecen fijos en la expresión los denominaremos **libres o fijos**.

Ejemplo:

1. Representar la formula $a^i \cdot b_i$ para $n=4$:

$$a^1 \cdot b^1 + a^2 \cdot b^2 + a^3 \cdot b^3 + a^4 \cdot b^4$$

2. Representar la formula $i \cdot a^i_{ik}$ para $n = 4$ e identificar cuáles son los índices libres de la expresión:

$$1 \cdot a^1_{1k} + 2 \cdot a^2_{2k} + 3 \cdot a^3_{3k} + 4 \cdot a^4_{4k}$$

Los índices libres son 'j' y 'k'.

- 3.** Aplicar el Convenio de Einstein con el siguiente sumatorio:

$$a_1 \cdot b_1 + a_2 \cdot b_2 + a_3 \cdot b_3 + a_3 \cdot b_3 +$$

La expresión que buscamos sería la siguiente:

$$a_j \cdot b_j \text{ con } n=3$$

- 4.** Simplificar lo máximo posible el siguiente sistema de ecuaciones:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 \\ y_2 = a_{21}x_1 + a_{22}x_2 \end{cases}$$

Tras una primera simplificación obtenemos:

$$\begin{cases} y_1 = a_{11}x_i \\ y_2 = a_{21}x_i \end{cases}$$

Y si observamos detenidamente, veremos que podemos aplicar el Convenio una vez más:

$$y_j = a_{jj}x_i$$

- 5.** Desarrollar la siguiente expresión a través del Convenio de Einstein:

$$\bar{T} = a_{ir}T_r \text{ para } n=4$$

Desarrollando el índice mudo 'r', tenemos lo siguiente:

$$\bar{T} = a_{i1}T_1 + a_{i2}T_2 + a_{i3}T_3 + a_{i4}T_4$$

Y haciendo uso del índice libre 'i' tenemos:

$$\bar{T}_1 = a_{11}T_1 + a_{12}T_2 + a_{13}T_3 + a_{14}T_4$$

$$\bar{T}_2 = a_{21}T_1 + a_{22}T_2 + a_{23}T_3 + a_{24}T_4$$

$$\bar{T}_3 = a_{31}T_1 + a_{32}T_2 + a_{33}T_3 + a_{34}T_4$$

$$\bar{T}_4 = a_{41}T_1 + a_{42}T_2 + a_{43}T_3 + a_{44}T_4$$

2.6.1. Concepto de rango y dimensión de un tensor

Denominamos **rango r** de un tensor u orden de un tensor al grado de libertad del mismo.

Denominamos **dimensión N** al espacio vectorial sobre el cual ubicamos el tensor.

Los tensores los podemos clasificar según su rango como:

- **Tensores de rango 0.** Son escalares, ya que son cantidades que tienen una magnitud, pero ninguna dirección asociada. Tienen un único componente.
 - Ejemplos: temperatura, presión, carga eléctrica, etc.

- **Tensores de rango 1.** Son vectores, ya que son cantidades que tienen magnitud y dirección asociadas. Tienen un grado de libertad y N componentes.
 - Ejemplos: velocidad, fuerza, peso, etc.
- **Tensores de rango 2.** Son matrices, ya que son cantidades que tienen magnitud y dos direcciones. Tienen dos grados de libertad y N componentes.
 - Ejemplo: tensión, conductividad eléctrica, inercia, etc.

La cantidad de combinaciones distintas que podemos hacer entre los rangos y las dimensiones de un tensor son:

$$n^o = r^N$$

Si a lo anteriormente expuesto, añadimos que cada componente del tensor puede tener a su vez 'n' variables, nos abre un mundo de posibles combinaciones:

- Rango = 0 y 0 variables en sus componentes → Escalar
- Rango = 1 y 0 variables en sus componentes → Vector
- Rango = 1 y 1 variables en sus componentes → Curvas
- Rango = 1 y 2 variables en sus componentes → Superficies
- Rango = 1 y 3 variables en sus componentes → Volúmenes
- Rango = 2 y 0 variables en sus componentes → Matrices

2.6.2. Tensores covariantes y contravariantes

Cuando definímos al principio el concepto de tensor, hablábamos de que consistía en una aplicación sobre un espacio. Llevado esto a una representación más formal tendríamos:

$$f: V_1 \times V_2 \times \dots \times V_n \rightarrow W$$

Donde tanto V como W son espacios vectoriales. Asumiendo que todas las V tienen la misma naturaleza y asumiendo que el espacio al que las vamos a llevar va a ser $W = \mathbb{R}$. Tenemos que:

n veces

$$T: V_x \times \dots \times V \rightarrow \mathbb{R}$$

Diremos que el tensor es n veces **covariante**.

Supongamos ahora que no todas las V tienen la misma naturaleza y que hay dos grupos bien diferenciados, teniendo un nuevo tensor tal que:

r veces

s veces

$$T: V_x^* \times \dots \times V_x^* \times V_x \times \dots \times V \rightarrow \mathbb{R}$$

Tendremos un tensor de tipo (r,s) que será r veces contravariante y s veces covariante.

Por nomenclatura, se suelen denominar como covariantes a los componentes con subíndices, mientras que a los superíndices se les llama contravariantes.



Ejemplo

$$T_i^j = a_i b_j$$

Donde “a” sería un componente covariante y “b” contravariante.

2.6.3. Introducción a TensorFlow

TensorFlow es una plataforma de código abierto de un extremo a otro para el aprendizaje automático. Tiene un ecosistema integral y flexible de herramientas, bibliotecas y recursos comunitarios que permite a los investigadores impulsar mejoras y avances en ML (*Machine Learning*) y a los desarrolladores crear e implementar fácilmente aplicaciones impulsadas por ML.

TensorFlow nos facilita librerías de código para poder probar y desarrollar, tanto en Python como JavaScript u otros lenguajes, redes neuronales que nos permitan encontrar patrones y adaptarlos a cada paso que demos con ellas.

Como ejemplo con esta herramienta, vamos a definir tensores que nos ayuden en un modelo de aprendizaje automático a través de diversas iteraciones para conseguir simular la siguiente función:

$$f(x) = 2x + 1$$

Haciendo uso de la versión de TensorFlow para JavaScript, crearemos un ejemplo en HTML para poder visualizarlo en el navegador sin necesidad de instalar nada:

```

1 <html>
2   <head>
3     <script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs/dist/tf.min.js"> </script>
4   </head>
5   <body>
6     <div id="output_field"></div>
7   </body>
8   <script>
9     async function learnLinear(){
10       const model = tf.sequential();
11       model.add(tf.layers.dense({units: 1, inputShape: [1]}));
12
13       model.compile({
14         loss: 'meanSquaredError',
15         optimizer: 'sgd'
16       });
17
18       const xs = tf.tensor2d([0, 1, 2, 3, 4], [5, 1]);
19       const ys = tf.tensor2d([1, 3, 5, 7, 9], [5, 1]);
20
21       await model.fit(xs, ys, {epochs: 300});
22
23       document.getElementById('output_field').innerText =
24         model.predict(tf.tensor2d([5, 6, 7], [3, 1]));
25
26     }
27     learnLinear();
28   </script>
29 </html>

```

Figura 51. Código de entrenamiento con TensorFlow. Fuente: elaboración propia.

Primeramente creamos un HTML lo más simple posible, con cualquier editor de texto y lo guardamos como un fichero html:

```

1 <html>
2   <head>
3     </head>
4   <body>
5     <div id="output_field"></div>
6   </body>
7 </html>

```

Figura 52. Código de una página HTML básico. Fuente: elaboración propia.

Como vemos, hemos creado una sección usando la etiqueta `<div>` a la que le hemos añadido el atributo `id` para poder identificar dicha sección.

A continuación vamos a añadir la librería de TensorFlow para JavaScript con una etiqueta `<script>`:

```
<script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs/dist/tf.min.js"></script>
```

Luego generamos un bloque para nuestro script usando la misma etiqueta de antes, solo que lo insertaremos después del body de nuestro HTML.

Primeramente, generaremos una función asíncrona para que esperemos a que termine el aprendizaje, a la que llamaremos `learnLinear`:

```

1  async function learnLinear() {
2
}

```

Figura 53. Función asíncrona en JavaScript. Fuente: elaboración propia.

Luego, vamos a generar nuestro modelo usando el método `sequential`, donde las salidas de una iteración se convierten en las entradas de la siguiente:

```
const model = tf.sequential();
```

Vamos a añadir a nuestro modelo una capa densa en la que todos los nodos de la red neuronal estarían conectados con todos. Puesto que vamos a usar solo un nodo, esto sería redundante.

Para ello, usaremos el método `dense` del atributo `layers` con la siguiente configuración para un único nodo:

```
model.add(tf.layers.dense({units: 1, inputShape: [1]}));
```

Ahora vamos a configurar algunos valores para el compilador del modelo. Para la función de pérdida, usaremos el error cuadrático medio y para el optimizador usaremos el gradiente descendente estocástico, o como sería en inglés stochastic gradient descent (SGD):

```

6   model.compile({
7     loss: 'meanSquaredError',
8     optimizer: 'sgd'
9   });

```

Figura 54. Configuración del compilador. Fuente: elaboración propia.

Los datos con los que vamos a entrenar nuestra red neuronal van a ser los que obtengamos de forma natural de la propia función. Llegados a este punto vamos a usar dos tensores, uno para los valores de x y otro para los resultados de $f(x)$ para dichos valores, tal que:

x	0	1	2	3	4
$f(x)$	1	3	5	7	9

```
11 const xs = tf.tensor2d([0, 1, 2, 3, 4], [5, 1]);
12 const ys = tf.tensor2d([1, 3, 5, 7, 9], [5, 1]);
```

Figura 55. Creación de tensores. Fuente: elaboración propia.

Donde el tensor queda definido por el método *tensor2d*, en el cual el primer parámetro es el conjunto de valores ya calculados previamente y el segundo parámetro es la definición de la matriz sobre la que se asienta, es decir, 5 filas y 1 columna.

Ahora vamos a entrenar nuestro modelo con el método *fit*, insertando los datos de entrada, los de salida y definiendo la cantidad de iteraciones que vamos a realizar (también llamadas “épocas”):

```
awaitmodel.fit(xs, ys, {epochs: 300});
```

Una vez hecho esto, ya estamos listos para poder usarlo, por ejemplo, con una predicción para un valor de entrada. Usaremos el método *predict* para hallar el valor que nuestro sistema ya entrenado predice para las entradas 5, 6 y 7. Dicho valor lo agregaremos como texto a la sección *div* que habíamos creado previamente usando el *Id* de la misma:

```
document.getElementById('output_field').innerText=
  model.predict(tf.tensor2d([5, 6, 7], [3, 1]));
```

Una vez hecho esto, podemos abrir nuestro fichero en un navegador donde obtendremos el siguiente resultado:

```
Tensor
[[11.0406675],
 [13.0595903],
 [15.0785122]]
```

Figura 56. Vista en el navegador de la salida. Fuente: elaboración propia.

Como vemos, tenemos una salida muy aproximada a lo que debería ser el resultado de la operación. Podemos aumentar el número de épocas (o iteraciones) para intentar conseguir más precisión.

Resumen

El álgebra es una rama de las matemáticas que permite trabajar con grandes cantidades de datos, facilitando, por tanto, la programación de las máquinas con las que estamos trabajando. Aunque se carezca de ciertas definiciones claras, a través de la intuición se puede llegar a entender claramente a qué se hace referencia en álgebra cuando se habla de conjuntos y elementos. Durante este tema se ha hablado no solo de

estos conceptos, sino también de las partes de los mismos y de las operaciones que se pueden llevar a cabo entre varios conjuntos de elementos.

En cuanto a las relaciones y aplicaciones que pueden darse en los conjuntos, se han revisado las nociones de par ordenado y de producto cartesiano, para poder más tarde hablar de las relaciones de equivalencia y de las aplicaciones inyectivas, sobreyectivas y biyectivas.

La combinatoria hace referencia a la forma en la que se realizan operaciones con los conjuntos y que se comenzó a abordar desde el concepto simple del principio de cardinalidad, que permite realizar operaciones sencillas basadas en los principios de adición, de Dirichlet, de inclusión-exclusión y de multiplicación. Más tarde, se habló de conceptos más complejos como son los de permutaciones (con y sin repetición), combinaciones (con y sin repetición) y variaciones (con y sin repetición).

En cuanto a los espacios vectoriales, se definió el espacio vectorial como tal, asociándolo, a su vez, con las propiedades del mismo y se realizó un acercamiento al concepto de combinación lineal. Se habló también de los subespacios lineales y de los sistemas libres, generadores y bases. Todos estos puntos han sido abordados a través de los teoremas que explican su comportamiento, así como su uso en IA.

Se ha realizado un acercamiento al concepto y trabajo con matrices, desde las operaciones a los principios que rigen la relación entre ellas. También se ha incluido el concepto de rango, determinante y matriz inversa como ineludibles para llegar a entender los hiperplanos reales.

Por último, se ha examinado también el concepto de cálculo tensorial, siendo este una generalización del álgebra lineal. En este apartado, se han desarrollado conceptos como el de tensor físico y el sistema de referencia al que se asocia; además, se han analizado los conceptos de rango y dimensión de un tensor. Dentro de los principios de trabajo con tensores, han surgido los conceptos de tensores covariantes y contravariantes.

En último lugar, se ha trabajado de forma más específica con el programa TensorFlow, plataforma de código libre, que permite el trabajo con *Deep Intelligence*.



Capítulo 3

Análisis

Introducción

Como se ha venido viendo hasta ahora, las matemáticas y sus partes son un elemento indispensable dentro de la Inteligencia Artificial o el *Machine Learning*. En este capítulo, se hablará del cálculo y el cálculo multivariante, también conocido como cálculo vectorial, ya que es imprescindible dominar el mismo para poder trabajar posteriormente con vectores e integradas, entre otras operaciones.

Por ello, durante este capítulo se trabajará con el cálculo infinitesimal y, concretamente, con los límites, derivadas, y las primitivas e integral definida, para terminar manejando los algoritmos de *gradient descent* y *forward & back propagation*.

En el último momento, además, se llevarán estos cálculos a la herramienta de código abierto TensorFlow, que ya se utilizó con anterioridad.

Objetivos

- Conocer los conceptos de cálculo relacionados con la Inteligencia Artificial.
- Realizar trabajos de cálculo matemático, como son los de límites y derivadas, entre otros.
- Aplicar los cálculos necesarios a través de la herramienta TensorFlow.

3.1. Cálculo infinitesimal

En este apartado, veremos de forma rápida los conceptos básicos del análisis real –funciones reales de variables reales y límites de funciones– para definir el concepto de derivada de función. Veremos también que las derivadas se utilizan para encontrar máximos y mínimos de funciones reales.

En aprendizaje automático (*machine learning*) e inteligencia artificial es muy común tener que encontrar mínimos y máximos de funciones reales. Por ejemplo, al entrenar cualquier modelo predictivo, el objetivo es casi siempre minimizar algún tipo de error de predicción (o de coste) o maximizar una función objetivo que representa el problema.

3.1.1. Límites

Definimos ahora el concepto de límite de una función real. Dada una función real f y un número $L \in \mathbb{R}$, decimos que el límite de $f(x)$ es L cuando x tiende a $c \in \mathbb{R}$. Si podemos siempre encontrar un valor x cerca de c tal que $f(x)$ esté tan cerca de L como se desee, escribimos:

$$\lim_{x \rightarrow c} f(x) = L$$

Formalmente, el límite de una función se puede determinar con la definición épsilon-delta. Decimos que $\lim_{x \rightarrow c} f(x) = L$ solo si para cualquier número $\epsilon > 0$ existe un número $\delta > 0$ tal que, si $|x - c| < \delta$, entonces $|f(x) - L| < \epsilon$.

Es importante recordar que el límite de una función en un punto dado puede no existir. En otras palabras, no existe ningún número L tal que el límite de la función sea igual a L .

Se dice que una función real es una **función continua** en el punto p si el límite de la función es igual al valor de la función $f(p)$:

$$\lim_{x \rightarrow p} f(x) = f(p)$$

Entonces, para que una función sea continua en un punto p , tienen que cumplirse las siguientes condiciones:

$f(p)$ existe, es decir, p pertenece al dominio de definición de f .

El límite $\lim_{x \rightarrow p} f(x)$ existe.

El límite es igual al valor de la función en el punto $\lim_{x \rightarrow p} f(x) = f(p)$.

Se dice que una función es continua si es continua en todos los puntos de su dominio. El conjunto de las funciones reales continuas se denota con $C(R)$.

3.1.1.1. Funciones

Una función (o mapa) $f : A \rightarrow B$ es una regla que asigna a cada elemento de A un único elemento del segundo conjunto B . A es el dominio de la función f o conjunto de definición. Por otra parte, el conjunto B de posibles valores que toma la función se llama codominio de f . Para cada valor $x \in A$, la salida o valor de retorno (*output*) de la función se denota con $f(x)$. La imagen inversa o preimagen de un subconjunto del codominio $C \subseteq B$, $f^{-1}(C)$ es el conjunto de puntos del dominio cuya imagen está en C :

$$f^{-1}(C) = \{x \in A \text{ tal que } f(x) \in C\}$$



Ejemplo

Sea f la función desde los números $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ en el conjunto de los meses del año. La función asigna a cada punto x su respectivo mes en el calendario, es decir, $f(1) = \text{enero}$, $f(2) = \text{febrero}$... El dominio de f es el conjunto de números naturales entre 1 y 12 (incluido los extremos) y el codominio es el conjunto de los meses del año.

Hablando formalmente, una función está definida no solo por la regla o las operaciones matemáticas que permiten calcular la salida (*output*) de la función para cada entrada (*input*), sino también por el dominio y el codominio. Una función con la misma regla del ejemplo precedente, pero con los números naturales entre 3 y 8 como dominio, no sería la misma función que la del ejemplo.

El subconjunto de puntos del codominio, que efectivamente son la salida de algunos valores del dominio, se llama **Imagen de la función** y se denota con $\text{Img}(f)$. Está claro que $\text{Img}(f) \subseteq B$, donde B es el codominio de la función.

En este apartado nos interesan las funciones de variable real a valores reales, es decir, cuando tanto el dominio como el codominio son subconjuntos de los números reales.



Ejemplo

Sea f la función desde $[1, 4]$ en $[1, 16]$, que asigna a cada punto x su cuadrado x^2 , es decir, $f(x) = x^2$. El dominio de f es el conjunto de puntos entre 1 y 4 (incluidos los extremos) y el codominio es el conjunto de puntos entre 1 y 16 (incluidos los extremos). Dado que tanto el dominio como el codominio son subconjuntos de los números reales, la función f es una función real de variable real.

Recordamos aquí la notación para los intervalos en los números reales:

El intervalo cerrado $[a,b]$ indica el conjunto de números reales entre los valores a y b , extremos incluidos, es decir, $[a,b] = \{x \in \mathbb{R} : a \leq x \leq b\}$.

Intervalo abierto: $(a,b) = \{x \in \mathbb{R} : a < x < b\}$.

Intervalo abierto a la izquierda y cerrado a la derecha: $(a,b] = \{x \in \mathbb{R} : a < x \leq b\}$.

Intervalo cerrado a la izquierda y abierto a la derecha: $[a,b) = \{x \in \mathbb{R} : a \leq x < b\}$.

En las funciones reales, muchas veces el dominio y el codominio no se especifican. En este caso, se toma como dominio de la función el subconjunto de los números reales, maximal con respecto a la inclusión, de modo que la función está bien definida. Por esta razón, el dominio de una función también se llama **conjunto de definición**. El codominio de la función siempre puede ser considerado el conjunto de todos los números reales \mathbb{R} .



Ejemplo

Sea $f(x) = \ln(x)$ la función logaritmo natural. El logaritmo natural solo es definido para números positivos y, entonces, podemos considerar como dominio de la función f el conjunto $\{x \in \mathbb{R} : x > 0\}$ o, en notación de intervalos, $(0, +\infty)$.

Sea $g(x) = \frac{1}{(x-5)}$. Dicha función se puede definir para todos $x \neq 5$, dado que para $x = 5$ el denominador de la fracción es 0. Entonces, como conjunto de definición de g , se puede considerar el conjunto de números reales $(-\infty, 5) \cup (5, +\infty)$.

3.1.1.2. Gráficas de funciones

Cada función real de variable real puede ser representada gráficamente mediante la **gráfica de función asociada**. Formalmente, dada una función real f , podemos definir la gráfica de la función como el subconjunto del plano cartesiano identificado por los puntos $(x, f(x))$ al variar $x \in A$ en el dominio de f .



Enlace de interés

Una forma muy rápida y sencilla de dibujar gráficas de funciones reales se puede encontrar en el buscador de Google. Basta con escribir la función que se quiere dibujar en el buscador y aparecerá la gráfica de la función. En la Figura 3 es posible ver el resultado de escribir la función $x^2 + x$ en el buscador.

www.google.com

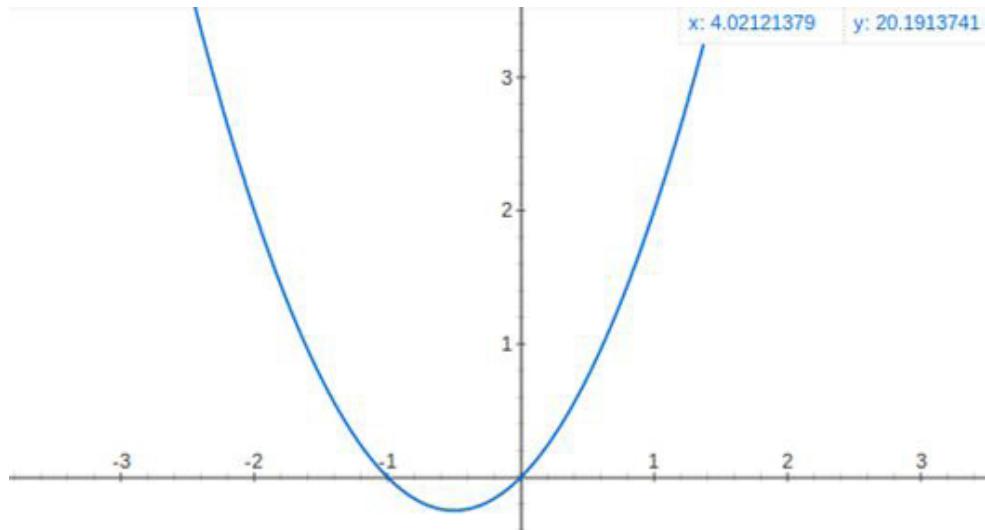


Figura 57. Gráfica de la función $f(x) = x^2 + x$.

3.1.2. Derivadas de funciones

La derivada de una función en un punto x es un valor numérico que representa cómo dicha función cambia al variar el valor de x . Geométricamente, se puede interpretar como el coeficiente de la recta tangente a la gráfica de la función en el punto $(x, f(x))$.

La derivada de una función f en x se define con el siguiente límite (si este existe):

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

Si la derivada existe, es decir, si el límite es definido en el punto x , decimos que la función f es derivable en x y denotamos con $f'(x)$ el valor de la derivada en el punto x . Existe otra notación para la operación de derivación:

$$f'(x) = \frac{df(x)}{dx}$$

Esta notación será útil más adelante, cuando consideraremos las funciones con más de una variable. En ese momento, tendremos que especificar respecto a qué variable se deriva la función.

Podemos observar que la derivada de una función es igual al límite de las pendientes de las rectas secantes de la gráfica de la función en los puntos $(x, f(x))$ y $(x + h, f(x + h))$. Como se puede ver en la Figura 4, el límite de estas rectas es exactamente la recta tangente en el punto $(x, f(x))$. Tras esta observación podemos concluir la primera regla para calcular derivadas, dado que la recta tangente a una línea recta es ella misma y que las líneas rectas son las gráficas de funciones del tipo $f(x) = ax + b$. Para estas funciones, $f'(x) = a$, es decir, que la derivada de una función lineal es una constante igual al coeficiente de la función. En el caso particular $f(x) = b$, obtenemos que la derivada de una función constante ($a = 0$) es 0.

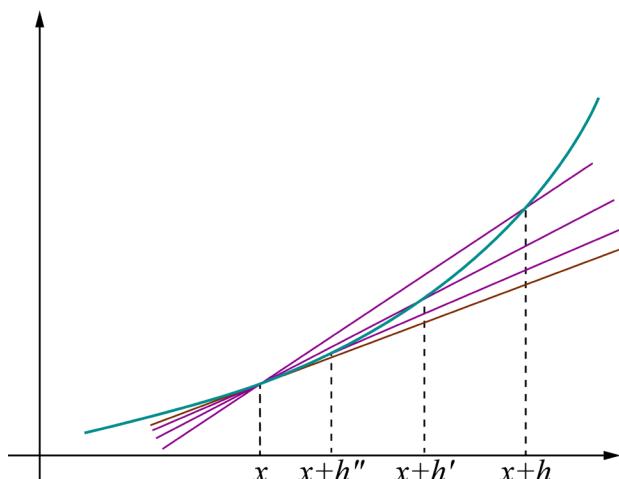


Figura 58. La recta tangente como límite de rectas secantes. Por Pbroks13 bajo licencia CC BY-SA 3.0. Disponible en <https://en.wikipedia.org/wiki/Derivative#/media/File:Lim-secant.svg>

Si la función es derivable para cada punto en un conjunto A , entonces se dice que la función es derivable en A . Está claro que la derivada f' es una función definida en el conjunto de puntos donde f es derivable. Además, es sencillo observar que, si una función es derivable en x , entonces tiene que ser continua en x . El hecho de que una función sea continua no es condición suficiente para su derivabilidad, sino que hay funciones que son continuas pero no derivables en algunos puntos.



Ejemplo

El ejemplo clásico de una función continua pero no derivable es la función valor absoluto.

Sea $f(x) = |x|$. Para cada $x > 0$, la función es igual a la función $f(x) = x$, cuya derivada es 1. De forma similar, para $x < 0$, la derivada es constante e igual a -1. En el punto $x = 0$, el límite $\lim_{h \rightarrow 0} \frac{f(h) - f(0)}{h}$ no existe, dado que para los valores de $h < 0$, el cociente es -1 y, para $h > 0$, el cociente es +1.

3.1.2.1. Calcular la derivada de una función

Veamos ahora las reglas básicas para calcular las derivadas de una función. En particular, veremos cuáles son las derivadas de algunas funciones básicas (polinomios, $\ln(x)$, e^x , $\sin(x)$, $\cos(x)$...) y cómo se calculan las derivadas de sumas, productos, fracciones y composiciones de funciones. De esta forma, es posible calcular las derivadas de todas las funciones que son combinaciones de funciones básicas.

A continuación se encuentran las derivadas de la mayoría de las funciones:

- **Monomios:** si $f(x) = x^n$, entonces $f'(x) = nx^{n-1}$.
- **Polinomios:** si $f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$, entonces $f'(x) = a_1 + 2a_2x^1 + 3a_3x^2 + \dots + na_nx^{n-1}$.
- **Exponenciales:** si $f(x) = e^x$, entonces $f'(x) = e^x$, es decir, la derivada de la función exponencial es la propia función exponencial.
- Logaritmos: si $f(x) = \ln(x)$, entonces $f'(x) = \frac{1}{x}$ para $x > 0$.
- Funciones trigonométricas:
 - Si $f(x) = \sin(x)$, entonces $f'(x) = \cos(x)$.
 - Si $f(x) = \cos(x)$, entonces $f'(x) = -\sin(x)$.
 - Si $f(x) = \tan(x)$, entonces $f'(x) = \frac{1}{\cos(x)^2} = 1 + \tan(x)^2$.
 - Si $f(x) = \cot(x)$, entonces $f'(x) = -\frac{-1}{\sin(x)^2} = -(1 + \cot(x)^2)$.
- Las derivadas precedentes se pueden combinar entre ellas de las siguientes formas:
 - Suma de funciones: si $f(x) = g(x) + h(x)$, entonces $f'(x) = g'(x) + h'(x)$.
 - Producto por escalar: si $f(x) = ag(x)$ para $a \in \mathbb{R}$, entonces $f'(x) = ag'(x)$.
 - Producto de dos funciones: si $f(x) = g(x)h(x)$, entonces $f'(x) = g'(x)h(x) + g(x)h'(x)$.
 - Cociente de dos funciones: si $f(x) = \frac{g(x)}{h(x)}$, entonces $f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{h(x)^2}$.
 - Composición de funciones: si $f(x) = g(h(x))$, entonces $f'(x) = g'(h(x))h'(x)$.



Ejemplo

Sea $f(x) = \exp(5x^3 - x + x^2 - 1)$. Aquí la derivada se puede calcular utilizando la regla de derivación para composición de funciones y las derivadas básicas para la función exponencial y para los polinomios. En particular, podemos escribir $f(x) = g(h(x))$, donde $h(x) = 5x^3 - x + x^2 - 1$ y $g(t) = \exp(t)$. Gracias a las reglas de derivación para polinomios y función exponencial, sabemos que $h'(x) = 15x^2 - 1 + 2x$ y que $g'(t) = e^t$. Ahora, utilizando la fórmula para derivar las funciones compuestas, obtenemos que:

$$f'(x) = \exp(5x^3 - x + x^2 - 1)(15x^2 - 1 + 2x)$$

Sea $g(x) = 2x^4 + \sin(3x)$. Para calcular la derivada de g , tenemos que sumar las derivadas de los dos términos de la suma. Entonces obtenemos $g'(x) = (2x^4)' + \cos(3x)(3x)' = 8x^3 + 3\cos(3x)$, donde la derivada de $\sin(3x)$ se obtiene aplicando la regla de derivación para la composición de las dos funciones $\sin(t)$ y $3x$.

3.1.3. Crecimiento, extremos relativos y extremos absolutos

Se dice que una función f definida en un conjunto (a,b) es una **función monótona creciente** si:

$$f(x) > f(y) \text{ para cada } x > y, x, y \in (a,b)$$

Se dice que una función f es una **función monótona decreciente** si:

$$f(x) < f(y) \text{ para cada } x > y, x, y \in (a,b)$$

Si una función es monótona creciente, entonces su gráfica será una línea creciente (Figura 5). Y si, por el contrario, una función es monótona decreciente, su gráfica será una línea decreciente al aumentar los valores de la variable x (Figura 6).

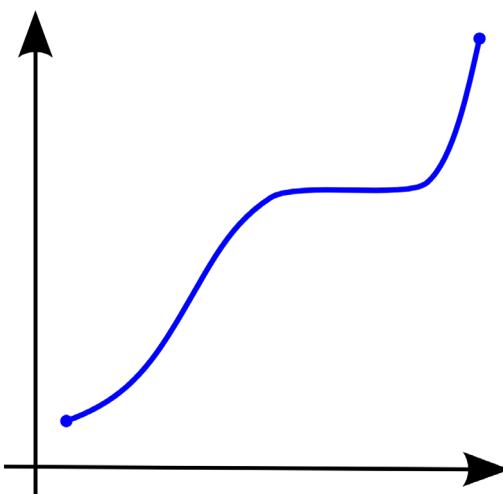


Figura 59. Gráfica de una función monótona creciente.

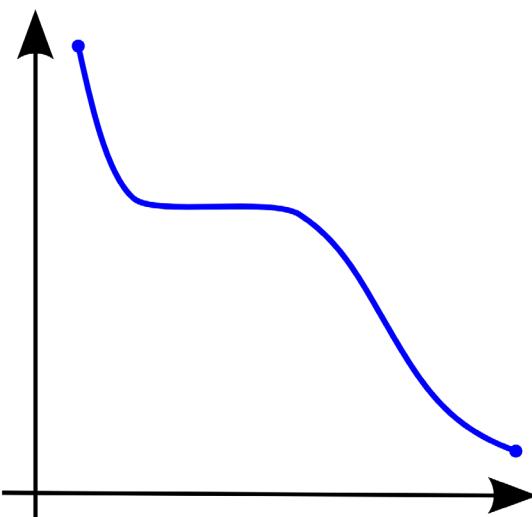


Figura 60. Gráfica de una función monótona decreciente.

Si consideramos las funciones derivables, entonces podemos relacionar el signo de la derivada con el comportamiento de las gráficas de las funciones. En particular, si la derivada es positiva en un punto x , entonces la recta tangente tiene pendiente positiva y la función es creciente alrededor del punto x . Por otro lado, si la derivada es negativa, entonces la función es decreciente alrededor del punto x .

De esta forma, comprobando el signo de la derivada $f'(x)$ de una función $f(x)$, es posible entender cómo la función crece o decrece al variar la variable x y dibujar, por lo tanto, una gráfica aproximada de $f(x)$.



Ejemplo

Sea $g(x) = \ln(x) + x^2$ para valores positivos. Su derivada es igual a $g'(x) = \frac{1}{x} + 2x = \frac{1+2x^2}{x}$ y es positiva para todos los valores en el dominio de la función. Así pues, podemos saber, sin tener que dibujar la gráfica con un ordenador y de forma más rigurosa, que la función g es monótona creciente para todos los valores en su dominio.



Ejemplo

Sea $h(x) = 4x - x^2$. Su derivada es claramente $h'(x) = 4 - 2x$ y, si observamos la desigualdad $h'(x) > 0$, vemos que la derivada es positiva para $x < 2$ y negativa para $x > 2$. Como se puede ver en la gráfica de la función en la Figura 7, $h(x)$ es creciente para $x < 2$ y decreciente para $x > 2$. El punto $x = 2$ (donde la derivada es 0) es el máximo de la función. Esta observación se puede extender a casos generales.

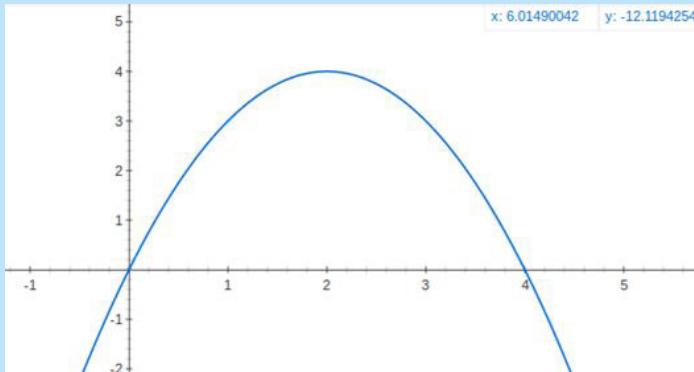


Figura 61. Gráfica de la función $h(x) = 4x - x^2$.

3.1.4. Máximos, mínimos y puntos críticos: optimización

Un punto de máximo, o más precisamente punto de máximo absoluto (o punto de máximo global) de una función f , definida en el dominio A , es cada punto $x_0 \in A$ tal que:

$$f(x_0) \geq f(x) \text{ para cada } x \in A$$

De forma opuesta, un punto de mínimo absoluto (o punto de mínimo global) es cada punto $x_0 \in A$ tal que:

$$f(x_0) \leq f(x) \text{ para cada } x \in A$$

Una función puede no tener máximos y/o mínimos absolutos. Si la función f no admite ningún máximo absoluto, entonces, desde la definición, obtenemos que para cada $x_0 \in A$ existe otro punto en el dominio $x \in A$ tal que $f(x) \geq f(x_0)$ y, entonces, eligiendo varios puntos en el dominio de la función podemos alcanzar valores siempre mayores. De forma similar, si la función no admite un mínimo absoluto, entonces podemos alcanzar valores siempre menores.



Ejemplo

Sea $f(x) = -x^2$. Está claro que la función f es menor o igual que 0 para cada $x \in \mathbb{R}$ y, además, $f(0) = 0$. Así pues, $x_0 = 0$ es el (único) punto de máximo absoluto de f . Se puede también observar que la función f no admite ningún mínimo absoluto en su dominio. De hecho, si hacemos crecer los valores de x , podemos alcanzar cualquier valor negativo.

Una función puede no admitir máximos y/o mínimos absolutos cuando estos valores están en el borde del dominio (que no es cerrado). Estos casos suelen ocurrir cuando el dominio de la función está restringido por razones ajenas a la propia función.



Ejemplo

Sea $h(x) = x$ definida en el intervalo abierto $(0, 2)$. Está claro que la gráfica de dicha función es una recta entre los puntos $(0, 0)$ y $(2, 2)$, pero estos dos puntos son excluidos por la definición del dominio de h como intervalo abierto. Entonces, h no admite ni máximo ni mínimo absoluto. Podemos comprobar este hecho para el mínimo de la siguiente manera: si $x_0 \in (0, 2)$, tenemos que $x = \frac{x_0}{2} \in (0, 2)$ y, además, $f(x) = x = \frac{x_0}{2} < x_0 = f(x_0)$. Por lo tanto, x_0 no puede ser punto de mínimo absoluto. Dado que el razonamiento se puede repetir para cualquier $x_0 \in (0, 2)$, hemos demostrado que h no admite mínimo en su dominio.

El teorema de Weierstrass, o teorema de los valores extremos, establece que una función continua definida en un intervalo cerrado y acotado (o una unión de intervalos cerrados y acotados) admite puntos de máximo y mínimo absolutos en dicho intervalo.

Otros tipos de puntos de máximo y mínimo son los **puntos de máximo o mínimo relativos o locales** (véase Figura 8). Un punto de máximo relativo o local de una función $f: A \rightarrow R$ es un punto x_0 tal que existe un entorno $B(x_0, \delta) = (x_0 - \delta, x_0 + \delta)$ y se cumple:

$$f(x_0) \geq f(x) \text{ para cada } x \in B(x_0, \delta) \cap A$$

Es decir, x_0 es punto de máximo absoluto en un entorno. De forma similar, se definen los puntos de mínimos relativos o locales. Observamos que todos los puntos extremos absolutos de una función son también extremos relativos. Si la función f es derivable, los puntos extremos coinciden con los puntos críticos de f .

Un **punto crítico** (o punto estacionario) de una función f es un punto x_0 perteneciente al dominio de f tal que la derivada f' es nula, es decir $f'(x_0) = 0$. Además, se suelen añadir a los puntos críticos también los puntos donde la función no es derivable y los extremos del dominio.

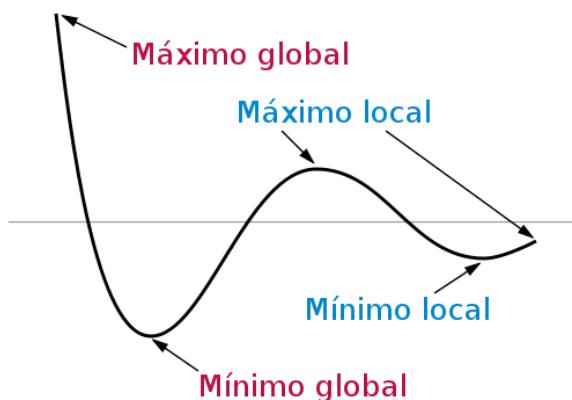


Figura 62. Máximos y mínimos de una función.

El teorema de Fermat expresa la conexión entre puntos críticos y extremos de una función. Si una función alcanza un máximo o mínimo relativo en el punto x_0 , y si la derivada existe en dicho punto x_0 , entonces $f'(x_0) = 0$. Dicho de otra forma, el teorema de Fermat afirma que ser punto estacionario es condición necesaria para ser punto máximo o mínimo local.

Ejemplos:

Sea $g(x) = x^3 - 6x^2 + 1$ una función polinómica y, por lo tanto, derivable en todos los valores. Sus puntos críticos se encuentran resolviendo la ecuación $g'(x) = 0$.

$$\begin{aligned} g'(x) &= 3x^2 - 12x = 0 \\ 3x(x - 4) &= 0 \end{aligned}$$

Así pues, las soluciones son $x = 0$ y $x = 4$, que son los dos puntos críticos de g . Para saber si los puntos críticos son puntos de máximo o mínimo locales hay que observar el signo de la derivada.

$$g'(x) = \begin{cases} g'(x) > 0 \text{ para } x < 0 \\ g'(x) < 0 \text{ para } 0 < x < 4 \\ g'(x) > 0 \text{ para } x > 4 \end{cases}$$

Así pues, está claro que $x_1 = 0$ es punto de máximo local y $x_2 = 4$ es punto de mínimo local.

Consideremos ahora un ejemplo de función no derivable. Sea $f(x) = |x| + x^3$. Esta función no es derivable en 0, dado que el valor absoluto no lo es. De todos modos, podemos calcular la derivada de forma usual para valores distintos de 0:

$$f'(x) = \begin{cases} -1 + 3x^2 \text{ para } x < 0 \\ 1 + 3x^2 \text{ para } x > 0 \end{cases}$$

Vemos que la derivada es siempre positiva para $x > 0$, y es 0 para $x_0 = -\sqrt[3]{3}$, cambiando de signo en este punto. En particular, $f'(x) < 0$ para $x_0 < x < 0$, y x_0 es punto de máximo local.

Dado que la función es continua en cada punto, y dado que la derivada cambia de signo en 0 (de negativa a positiva), 0 es punto de mínimo local. Hay que fijarse en que en 0 la derivada no está definida y este punto no se puede encontrar a partir de la ecuación $f'(x) = 0$.

3.1.4.1. Derivadas de órdenes superiores

Dado que la derivada de una función es una función real, podemos derivar otra vez la derivada de una función (y sucesivamente derivar una y otra vez).

De esta forma se definen las derivadas de una función de orden superior, de modo que podemos considerar la derivada de orden 2 o derivada segunda $f''(x) = (f'(x))'$, la derivada de orden 3. En general, se puede definir de forma iterativa la derivada de orden n :

$$f^n(x) = (f^{n-1}(x))' = \frac{d^n f(x)}{dx^n}$$

Donde los exponentes indican el grado de la derivada.

La derivada segunda es útil en la búsqueda de los puntos de máximo y mínimo locales. En particular, a través del signo de la derivada segunda en el punto crítico, es posible diferenciar entre máximo y mínimo. Si $f''(x) = 0$, es decir, si x_0 es punto crítico, entonces:

- Si $f''(x) > 0$, el punto x_0 es un punto de mínimo local.
- Si $f''(x) < 0$, el punto x_0 es un punto de máximo local.



Ejemplo

Si $h(x) = \sin(x)\cos(x)$, la derivada prima es:

$$h'(x) = \cos(x)^2 - \sin(x)^2$$

Por lo tanto, los puntos críticos se obtienen resolviendo la siguiente ecuación:

$$\begin{aligned} \cos(x)^2 &= \sin(x)^2 \\ \cos(x) &= \pm \sin(x) \end{aligned}$$

La ecuación anterior nos proporciona el conjunto de soluciones:

$$x = k \frac{\pi}{4} \quad k = 0, \pm 1, \pm 2, \dots$$

Utilizamos ahora la derivada segunda para identificar los máximos y los mínimos:

$$h''(x) = -2\cos(x)\sin(x) - 2\sin(x)\cos(x) = -4\sin(x)\cos(x)$$

Observamos ahora que $\cos(\pm \frac{\pi}{4} + 2k\pi)$ son valores positivos y $\cos(\pm 3\frac{\pi}{4} + 2k\pi)$ son valores negativos, mientras que, para la función \sin , obtenemos que $\sin(\frac{\pi}{4} + 2k\pi) > 0$, $\sin(3\frac{\pi}{4} + 2k\pi) > 0$ y $\sin(-\frac{\pi}{4} + 2k\pi) < 0$, $\sin(-3\frac{\pi}{4} + 2k\pi) < 0$. Juntando estas desigualdades obtenemos:

- $h''(\frac{\pi}{4} + 2k\pi) = -4\sin(\frac{\pi}{4} + 2k\pi)\cos(\frac{\pi}{4} + 2k\pi) < 0$ y entonces $\frac{\pi}{4} + 2k\pi$ son puntos de máximo.
- $h''(-\frac{\pi}{4} + 2k\pi) = -4\sin(-\frac{\pi}{4} + 2k\pi)\cos(-\frac{\pi}{4} + 2k\pi) > 0$ y entonces $-\frac{\pi}{4} + 2k\pi$ son puntos de mínimo.

De forma similar podemos obtener los resultados para los otros dos conjuntos de puntos críticos.

3.1.5. Primitivas e integral definida

3.1.5.1. Integrales de funciones

Existen varias definiciones de la integral de una función real, por lo que explicaremos ahora la más intuitiva: la integración de Riemann.

Sea f una función definida sobre un intervalo acotado $[a, b]$. Supongamos también que la función es positiva, es decir $f(x) \geq 0$ para cada $x \in [a, b]$.

Consideremos ahora el problema de calcular el área entre la gráfica de la función y y el eje de las x (Figura 11).

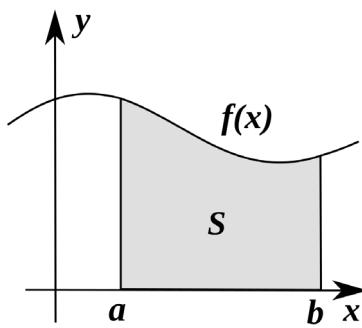


Figura 63. El área entre la gráfica de la función y el eje de las x .

Una primera aproximación se puede construir dividiendo el intervalo $[a, b]$ mediante una partición, es decir, una secuencia de números $a = x_0 < x_1 < x_2 < \dots < x_{(n-1)} < x_n = b$, y construir una aproximación del área mediante rectángulos (Figura 12). Escribimos la siguiente fórmula:

$$\text{Área} \approx \sum f(x_i) (x_i - x_{i-1}) = S(n)$$

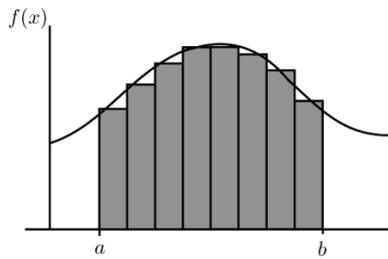


Figura 64. Aproximación del área bajo la gráfica de la función.

Intuitivamente, si aumentamos el número de puntos en la partición, la aproximación obtenida con $S(n)$ es cada vez más precisa y, en el límite para $n \rightarrow +\infty$, el valor de la suma (llamada suma de Riemann) es igual al área bajo la función.

Si existe el límite de las sumas de Riemann, este límite se llama integral y se denota de la siguiente forma:

$$\int f(x) dx$$

Donde se indican los extremos de integración a y b , y se explica con el símbolo dx cuál es la variable que nos interesa integrar.

Los límites de integración pueden también ser $a = -\infty$ y/o $b = +\infty$. Se puede, además, escribir el conjunto de integración debajo del símbolo de integral:

$$\int_a^b f(x) dx$$

Para las funciones no positivas, la integral se define descomponiendo las funciones en partes positivas y partes negativas:

$$\int f(x) dx = \int f^+(x) dx - \int f^-(x) dx$$

Donde $f^+(x) = \max\{f(x), 0\}$ y $f^-(x) = \max\{-f(x), 0\}$ son la parte positiva y negativa de la función.

A partir de la integral de una función, podemos definir una nueva función, llamada **integral indefinida**:

$$F(x) = \int f(t) dt$$

3.1.5.2. Primitivas de funciones y teorema fundamental del cálculo

La primitiva de una función es la inversa de la derivada, es decir, que para una función f la primitiva es una función $F(x)$ tal que $F'(x) = f(x)$.

Observamos que, partiendo del hecho de que la derivada de una función constante es igual a 0, obtenemos que la primitiva de una función no es única y, si F es una primitiva de f , entonces también $F + a$ lo es para cada $a \in \mathbb{R}$.

El teorema fundamental del cálculo dice que la integral indefinida de una función es una de sus primitivas:

$$\text{Si } F(x) = \int f(t) dt, \text{ entonces } F'(x) = f(x)$$

Además, nos permite calcular las integrales definidas con la siguiente fórmula:

$$\int f(x) dx = F(b) - F(a), \text{ donde } F \text{ es una primitiva de } f$$



Ejemplo

Una primitiva de la función $g(x) = x^2$ es $G(x) = \frac{x^3}{3}$. Es fácil de comprobar que:

$$G'(x) = \left(\frac{x^3}{3} \right)' = 3 \frac{x^2}{3} = x^2 = f(x)$$

Así que ahora podemos calcular cualquier integral del tipo $\int f(x) dx$.

$$\int x^2 dx = G(2) - G(0) = \frac{8}{3} - 0 = \frac{8}{3}$$

$$\int x^2 dx = \left(\frac{x^3}{3} \right)_{-1}^2 = \frac{8}{3} - \left(\frac{-1}{3} \right) = \frac{9}{3} = 3$$

Propiedades de la integral y algunas primitivas de funciones básicas

La primera propiedad es que la integral es una operación lineal:

$$\int f(x) + g(x) dx = \int f(x) dx + \int g(x) dx$$

$$\int f(x) dx = a \int f(x) dx \text{ para cada } a \in \mathbb{R}$$

No hemos especificado los intervalos de integración porque esta propiedad es válida para cualquier conjunto de integración y también para la integral indefinida.

La integral es también lineal en el conjunto de integración:

$$\int f(x) dx = \int f(x) dx + \int f(x) dx \text{ para cada } a \leq c \leq b$$

Y, en particular, es lineal para cada función f :

$$\int f(x) dx = 0$$

Es posible también definir la integral para intervalos invertidos mediante la siguiente fórmula:

$$\int f(x) dx = - \int f(x) dx$$

Muchas veces es importante acotar el valor de una integral, y para eso existen varias desigualdades que se pueden emplear. Las más básicas son las siguientes:

Si $f(x) \leq g(x)$ para cada $x \in [a, b]$, entonces $\int f(x) dx \leq \int g(x) dx$.

En particular, $\int f(x) dx \leq \int |f(x)| dx$.

Si $m = \min_{x \in [a, b]} f(x)$ y $M = \max_{x \in [a, b]} f(x)$, entonces $m(b - a) \leq \int f(x) dx \leq M(b - a)$.

Para calcular las primitivas de las funciones básicas, es suficiente invertir las tablas de las derivadas comunes.

Monomios: si $f(x) = x^k$, entonces una primitiva es $F(x) = \frac{x^{k+1}}{k+1}$.

Función exponencial: si $f(x) = e^x$, entonces $F(x) = e^x$.

Si $f(x) = \frac{1}{x}$, entonces $F(x) = \log(x)$ para $x > 0$.

...

Derivar una función es una operación simple y, si se conocen las reglas fundamentales, se reduce a un cálculo mecánico. Calcular una integral hallando la primitiva de la función es, en general, más difícil. Existen varias técnicas y trucos, pero también existen casos en los que no es posible encontrar una primitiva de la función expresa como combinación de funciones básicas, es decir, en fórmula cerrada. En muchos casos, es entonces necesario utilizar técnicas de aproximación numéricicas para calcular la integral de la función. Las técnicas más simples se basan en la suma de Riemann y algunas mejorías, y otras técnicas se basan en la probabilidad (método Montecarlo).

3.2. Cálculo multivariable

Consideremos ahora funciones con más de una variable real a valores en los números reales, es decir, una función como la siguiente:

$$f: A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$$

Para las funciones de dos variables, es posible dibujar las gráficas asociadas en tres dimensiones. Sin embargo, si hay más de dos variables, no es posible dibujar la gráfica de la función, dado que la gráfica de una función de n variables es un subconjunto de un espacio $n + 1$ -dimensional.



Ejemplo

Sea $g(x,y)$ la función de dos variables definida por $g(x,y) = x^2 + e^y$. La gráfica obtenida a través del buscador de Google se puede ver en la figura 9.

La función g en el punto $(0,0)$ toma el valor $g(0,0) = 0^2 + e^0 = 0 + 1 = 1$ y en el punto $(2, \ln(3))$ el valor $g(2, \ln(3)) = 2^2 + e^{\ln(3)} = 4 + 3 = 7$. Es fácil comprobar que la función es definida y positiva para todos valores de $(x,y) \in \mathbb{R}^2$.

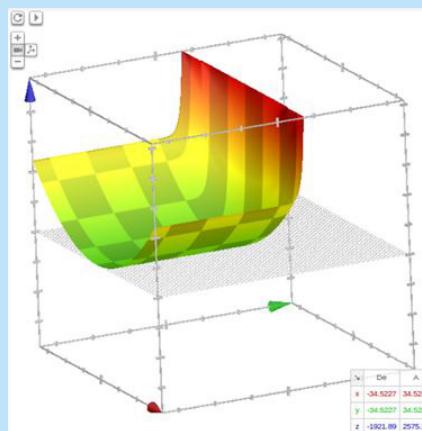
Ejemplo


Figura 65. Gráfica de la función $g(x,y) = x^2 + e^y$

3.2.1. Derivadas parciales

También las funciones de diversas variables se pueden derivar, pero, a diferencia de las funciones de una sola variable, no hay una sola derivada. Para cada variable, es posible definir una derivada prima de la función mientras se fijan el resto de las variables. En particular, si $f(x_1, x_2, \dots, x_n) = f(x)$ es una función de n variables, entonces la derivada parcial respecto a la variable x_i se denota así:

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h \rightarrow 0} \frac{f(x_1, \dots, x_2 + h, \dots, x_n) - f(x_1, \dots, x_n)}{h}$$

A veces, el símbolo f_{x_i} se puede utilizar en vez de $\frac{\partial f}{\partial x_i}$ para la derivada parcial de f respecto a la variable x_i . Para calcular las derivadas parciales de una función, se utilizan las mismas reglas que en el caso de una sola variable y se consideran las demás variables como constantes.

Ejemplo

Si $h(x_1, x_2, x_3) = x_1^3 x_2 + x_1 \ln(x_2) + 4x_3$, podemos calcular las tres derivadas parciales respecto a las distintas variables con las mismas reglas de las derivadas en el caso de las funciones de una sola variable.

$$\frac{\partial h}{\partial x_1}(x_1, x_2, x_3) = \frac{\partial (x_1^3 x_2)}{\partial x_1} + \frac{\partial (x_1 \ln(x_2))}{\partial x_1} + \frac{\partial (4x_3)}{\partial x_1} = 3x_1^2 x_2 + \ln(x_2) + 0$$

$$\frac{\partial h}{\partial x_2}(x_1, x_2, x_3) = \frac{\partial (x_1^3 x_2)}{\partial x_2} + \frac{\partial (x_1 \ln(x_2))}{\partial x_2} + \frac{\partial (4x_3)}{\partial x_2} = x_1^3 + \frac{x_1}{x_2} + 0$$

$$\frac{\partial h}{\partial x_3}(x_1, x_2, x_3) = \frac{\partial (x_1^3 x_2)}{\partial x_3} + \frac{\partial (x_1 \ln(x_2))}{\partial x_3} + \frac{\partial (4x_3)}{\partial x_3} = 0 + 0 + 4$$

Por ejemplo, para calcular la derivada $\frac{\partial(x_1^3 x_2)}{\partial x_1}$, consideramos x_2 como una constante y derivamos respecto a x_1 . La constante x_2 se puede sacar de la derivación y nos queda $x_2 \frac{\partial(x_1^3)}{\partial x_1}$, que es ahora una derivada usual de un monomio, de modo que obtenemos $x_2^2 (3x_1^2)$.



Ejemplo

Si $f(x,y) = \sin(xy)$ y calculamos la derivada parcial respecto a x , tenemos que considerar y como una constante y aplicar las reglas de las derivaciones.

$$\frac{\partial f}{\partial x}(x,y) = \frac{\partial \sin(xy)}{\partial x} = y \cos(xy)$$

3.2.2. Gradiente y derivadas direccionales

El gradiente de una función de n variables reales es el vector de dimensión n cuyos componentes son las derivadas parciales de la función respecto a las distintas variables. Formalmente, el gradiente de una función $f(x_1, \dots, x_n)$ se indica con el símbolo $\nabla f(x_1, \dots, x_n)$ y es el vector de derivadas parciales:

$$\nabla f = \begin{pmatrix} \frac{\partial h}{\partial x_1} \\ \frac{\partial h}{\partial x_2} \\ \vdots \\ \frac{\partial h}{\partial x_n} \end{pmatrix}$$

La derivada direccional de una función $f: A \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ es la derivada a lo largo de una dirección fija en el espacio vectorial de las variables. En particular, se define para cada vector $v \in A \subseteq \mathbb{R}^n$ en el dominio de la función:

$$\frac{\partial f}{\partial v}(x) = \lim_{h \rightarrow 0} \frac{f(x + hv) - f(x)}{h}$$

La derivada direccional se calcula a través del gradiente de la función con la siguiente fórmula:

$$\frac{\partial f}{\partial v}(x) = \nabla f(x) \cdot v$$

Donde el producto es un producto escalar entre vectores en \mathbb{R}^n .



Ejemplo

Consideremos la función de dos variables $g(x,y) = 2xy + y^2$ y veamos cómo calcular la derivada direccional en la dirección $v = (1, 2)$. Primero, calculamos el gradiente de la función ∇g :

$$\nabla g = \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \end{pmatrix} = \begin{pmatrix} 2y \\ 2x + 2y \end{pmatrix}$$



Ejemplo

Ahora podemos calcular la derivada direccional de la función a través del producto escalar entre el gradiente y la dirección:

$$\frac{\partial g}{\partial v} (x, y) = \nabla g(x, y) \cdot v = (2y, 2x + 2y) \cdot (1, 2) = 2y + 2(2x + 2y) = 4x + 6y$$

El gradiente de una función f en un punto $x \in R^n$ indica también la dirección que maximiza la derivada direccional de la función. Se puede demostrar que para cada vector:

$$\|v\|_2 = \max_{w \in R^n, \|w\|_2=1} v \cdot w = v \cdot \frac{v}{\|v\|_2}$$

Así pues, para el vector gradiente $\nabla f(x)$, la dirección que maximiza el producto escalar es exactamente la dirección del gradiente.

3.2.3. Segundas derivadas parciales y matriz hessiana

Igual que en el caso de las funciones de una sola variable, es posible definir las derivadas parciales de orden mayor que 1; la diferencia es que para las funciones de diversas variables podemos mezclar las derivadas. En particular, para las derivadas parciales de orden 2 obtenemos todas las derivadas de la forma $\frac{\partial^2 f}{\partial x_i \partial x_j}$.

Si $i = j$, escribimos simplemente $\frac{\partial^2 f}{\partial x_i^2}$. Además, el teorema de Schwarz afirma que, bajo hipótesis de regularidad (es decir, las derivadas segundas existen y son continuas), el orden de las derivadas parciales no influye:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Las derivadas segundas parciales definen, pues, una matriz simétrica de dimensiones $n \times n$, que se llama matriz hessiana:

$$H(f) = \begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{vmatrix}$$

3.2.4. Máximos, mínimos y puntos críticos: optimización

Los puntos de máximo y mínimo (locales y globales) de una función $f: R^n \rightarrow R$ se definen exactamente como para las funciones de una variable real.

Los puntos críticos de una función de diversas variables son los puntos tales que el vector gradiente es igual al vector de ceros. Así pues, $x = (x_1, \dots, x_n)$ es un punto crítico si:

$$\nabla f(x_1, \dots, x_n) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Es decir, $\frac{\partial f}{\partial x_i}(x_1, \dots, x_n)$ para cada variable x_i .



Ejemplo

Consideremos la función de dos variables $g(x,y) = e^{-x^2-y^2}$ y calculemos sus puntos críticos (a partir de la gráfica de la función en la figura 10, entendemos que debería haber un solo punto crítico).

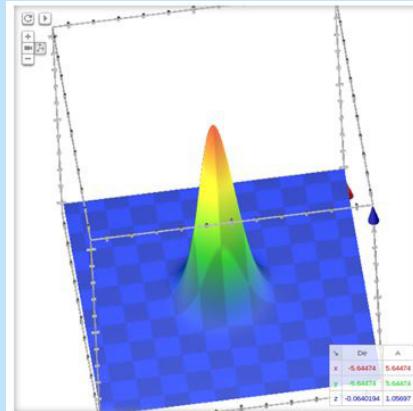


Figura 66. Gráfica de la función $g(x,y) = e^{-x^2-y^2}$.

El gradiente de la función es igual a lo siguiente:

$$\nabla g(x,y) = \begin{vmatrix} \frac{\partial g}{\partial x}(x,y) \\ \frac{\partial g}{\partial y}(x,y) \end{vmatrix} = \begin{pmatrix} -2xe^{-x^2-y^2} \\ -2ye^{-x^2-y^2} \end{pmatrix}$$

Está claro entonces que el único punto crítico es el punto $(0, 0)$.

3.3. Algoritmos de gradient descent y forward & backpropagation

Como hemos visto anteriormente los puntos críticos, junto con el cálculo de máximos y mínimos, juegan un papel muy importante en la optimización de funciones. Precisamente en base a esta optimización, vamos a ver a continuación un método que, a día de hoy, tiene un uso bastante extendido en el aprendizaje automático o ML (Machine Learning), y una forma para entrenar dichas redes neuronales a partir de él.

3.3.1. Gradient descent

Antes de ver en qué consiste el método del gradiente descendente, vamos a repasar un concepto previo.

Una forma de reconocer qué error se comete en cualquier modelo o sistema que pretenda asemejarse a otro es comparar el valor que obtenemos a la salida a partir de una entrada de datos en el sistema original SO, con el valor obtenido a la salida para el mismo conjunto de datos de entrada en el sistema a moldear SM. Esto lo podemos expresar mediante el error cuadrático medio o MSE (Mean Squared Error):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\text{SO}_i - \text{SM}_i)^2$$

Esto nos sirve para definir la función "coste" como:

$$E = \text{MSE}$$

Una vez definida nuestra función de coste, entendemos que lo que queremos es minimizar el error de la mejor forma posible. Para ello, vamos a hacer uso del gradiente, que, como recordaremos, consiste en un vector de derivadas parciales con el cual hallar puntos críticos como el mínimo de una función. Siguiendo con nuestra función de coste, vamos a suponer que posee un parámetro w tal que para $f(w)$ sea un mínimo de la función; entonces, las derivadas parciales serían:

$$\nabla f(w) = \left(\frac{\partial f}{\partial x_1}(w), \dots, \frac{\partial f}{\partial x_n}(w) \right)$$

Con lo que para averiguar para qué valores de w conseguimos optimizar la función $f(w)$ teniendo en cuenta que se trata de E , tendremos que hacer una serie de iteraciones e ir aplicando la siguiente regla en cada una de ellas:

$$w_{t+1} = w_t - \eta_t \nabla E(w_t)$$

O pasándolo a derivadas parciales:

$$w_{t+1} = w_t - \eta_t \frac{\partial E(w_t)}{\partial w}$$

A destacar lo siguiente:

- Usamos la resta para ir en dirección negativa al gradiente.
- El cálculo de la siguiente w se hace de forma proporcional al gradiente.
- El concepto η es un modulador del grado con el que calculamos el siguiente término de $\nabla E(w)$, también llamado **tasa de aprendizaje**.

Este último concepto de tasa de aprendizaje es muy importante:

- Para empezar $\eta_t > 0$.
- Si la tasa de aprendizaje es muy pequeña, el gradiente descendente tendrá que pasar por muchas iteraciones antes de converger a una solución válida y llevará mucho tiempo.
- Si la tasa de aprendizaje es muy grande, el gradiente descendente es posible que se salte el mínimo global e incluso termine en otro lado con una desviación aún mayor que la anterior, lo que podría hacer que no se encontrase una solución válida.

3.3.2. Forward & backpropagation

Llamamos propagación adelante y hacia atrás de errores (*forward & back propagation*) a una forma de utilizar el método anterior del gradiente descendente para entrenar redes neuronales. Consiste en ciclos en los cuales, a partir de una entrada, se propaga a través de las diversas capas neuronales hasta obtener una salida y, luego, al comparar esta con la salida deseada, obtendremos un error que propagaremos hacia atrás para recalcular todas las capas de neuronas que hayan tenido que ver en el proceso.

Veamos cómo sería la topología más habitual de un sistema neuronal:

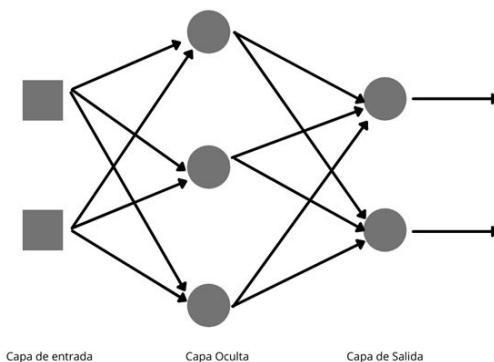


Figura 66. Ejemplo de red neuronal. Fuente: elaboración propia.

Como vemos, no hay realimentación y todos los nodos de una capa están conectados con todos los nodos de la siguiente. La organización por capas suele ser:

- Capa de entrada
- Capa oculta
- Capa de salida

Las funciones que definirían cada neurona son conocidas como funciones de activación y suelen tener la particularidad de ser funciones cuyas derivadas sean sencillas para poder ahorrar en coste computacional. Un ejemplo sería la función sigmoidal:

$$f(x) = \frac{1}{1 - e^{-x}}$$

Que, además, tiene la particularidad de que su derivada se puede expresar en función de sí misma:

$$\frac{df(x)}{dx} = f(x)(1 - f(x))$$

Ahora vamos a dar algunos consejos para mejorar la eficiencia del entrenamiento de nuestra red neuronal:

- Mezclar los ejemplos de forma que nunca sean consecutivos.
- Entrenar con ejemplos que puedan inducir al error con mayor frecuencia que otros que no lo hagan.
- La media de cada variable de entrada debería ser cero, de forma que tengamos los datos más estandarizados. A esto se le llama "desplazamiento".
- La varianza de las variables de entrada debería ser 1, de forma que vayan del $[-1, 1]$. A esto se le llama "escalado".
- Para evitar problemas en los sistemas que intenten realizar una clasificación, se recomienda que las salidas sea 0 -1 o 1 → $\{-1, 1\}$.

3.3.3. Aplicación y ejemplos con TensorFlow

Gracias a TensorFlow, el uso del gradiente es tan sencillo como poner la etiqueta ‘SGD’ como optimizador y la etiqueta ‘meanSquaredError’ al compilar nuestro modelo, cuando llamemos al método *compile*. Para el entrenamiento de la red neuronal, nos basta con invocar al método asociado a nuestro modelo *fit*, junto a los datos de entrada, salida y el número de épocas o ciclos del entrenamiento.

Igualmente podemos entrenar nuestro propio optimizador a partir de uno ya existente en la librería:

```
paso = tf.train.GradientDescentOptimizer(ta).minimize(cost)
```

Donde *cost* será nuestra función de coste y *ta* será la tasa o ratio de aprendizaje. Lo que nos devuelve es el paso de entrenamiento para poder empezar con las iteraciones.

Ahora vamos a ver de nuevo un ejemplo muy sencillo de red neuronal para ver cómo afecta de forma práctica el modelo de la misma, alterando las capas de las que está compuesta. Para ello vamos a usar la función:

$$f(x) = -x + 3$$

Partamos de la base de del siguiente código:

```

1  <html>
2    <head>
3      <script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs/dist/tf.min.js"></script>
4    </head>
5    <body>
6      <div id="output_field"></div>
7    </body>
8    <script>
9      async function learnLinear(){
10
11        const model = tf.sequential();
12        model.add(tf.layers.dense({units: 1, inputShape: [1]}));
13        model.add(tf.layers.dense({units: 1}));
14
15        model.compile({
16          loss: 'meanSquaredError',
17          optimizer: 'sgd'
18        });
19
20        const xs = tf.tensor2d([-2, -1, 0, 1, 2], [5, 1]);
21        const ys = tf.tensor2d([5, 4, 3, 2, 1], [5, 1]);
22
23        await model.fit(xs, ys, { epochs: 100,
24          callbacks: {
25            onEpochEnd: async(epoch, logs) =>{
26              console.log("Epoch:" + (epoch+1) + " Loss:" + logs.loss);
27            }
28          }
29        });
30
31        document.getElementById('output_field').innerText=
32          model.predict(tf.tensor2d([4], [1, 1]));
33
34      }
35      learnLinear();
36    </script>
37  </html>
```

Figura 67. Código base con TensorFlow. Fuente: elaboración propia.

Para empezar tenemos una red neuronal con una entrada, una capa oculta con un único nodo y una salida, lo que quedaría así:

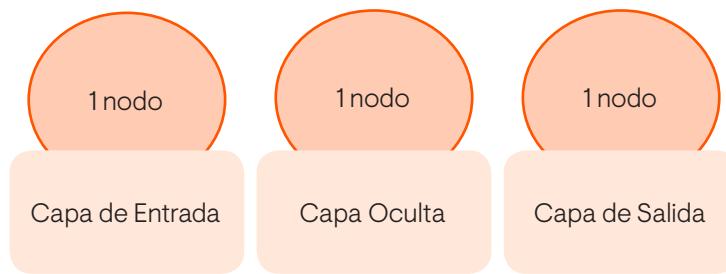


Figura 68. Vista red neuronal. Fuente: elaboración propia.

Para este ejemplo, cuando terminamos de hacer 100 épocas o ciclos de iteración, tenemos por consola lo siguiente:

Epoch:100 Loss:1.480785272178764e-9

Lo que nos viene a dar una idea de lo que se ha acercado nuestro sistema al mínimo absoluto de su desarrollo. Para la salida por pantalla, teniendo en cuenta que queríamos la salida para un valor de entrada de 4:

Tensor
[[-1.0002255],]

Figura 69. Tensor de salida para 100 épocas y un nodo en la capa oculta. Fuente: elaboración propia.

Ahora probemos qué pasaría si subiéramos a 200 el número de épocas del entrenamiento de nuestro sistema:

Epoch:200 Loss:0.000018080407244269736

Con una salida de:

Tensor
[[-1.0074611],]

Figura 70. Tensor de salida para 200 épocas y un nodo en la capa oculta. Fuente: elaboración propia.

Vemos que, aunque aumenta el número de épocas, el valor de pérdida, en la práctica, está cerca de no variar. Sin embargo, el valor que se nos muestra por pantalla del tensor, aunque cercano al real, ha empeorado con respecto al valor que teníamos antes, con tan solo 100 iteraciones.

Volvamos al valor de 100 de las épocas, pero esta vez vamos a variar la capa oculta, añadiendo tres nodos más para tener un total de cuatro, con lo que nos quedaría la siguiente red neuronal:



Figura 71. Vista red neuronal actualizada. Fuente: elaboración propia.

Para hacer esto, cambiaremos el modelo de forma que quede así:

```
constmodel = tf.Sequential();  
  
model.add(tf.layers.dense({units: 4, inputShape: [1]}));  
  
model.add(tf.layers.dense({units: 1}));
```

Cuando termina el entrenamiento tenemos:

```
Epoch:100 Loss:7.958675385388858e-10
```

Un valor de pérdida sumamente pequeño, y además:

```
Tensor  
[[-1.0000346],]
```

Figura 72. Tensor de salida para 100 épocas y 4 nodos en la capa oculta. Fuente: elaboración propia.

Tenemos un valor estimado mucho más cercano al real que hasta ahora. De hecho, si aumentásemos el valor del número de épocas, por ejemplo a 400, veríamos que ya no notaríamos diferencia en cuanto al valor de perdida y que la mejora en cuanto al valor estimado es escasa.

Resumen

En matemáticas, se define cálculo como aquella área que se encarga del estudio de la variación y movimiento, permitiendo prever el resultado que va a obtenerse de estos. El cálculo, como operación relacionada con la lógica, está ligado a la Inteligencia Artificial de manera indisoluble.

En este capítulo, se comienza hablando de los conceptos básicos de análisis real, funciones reales de variables reales y límites funcionales, como punto de partida que puede llevar a entender y trabajar de forma correcta con las derivadas y límites.

Se ha abordado también el concepto de función y, posteriormente, el de gráficas de funciones.

En segundo lugar, se ha examinado el concepto de derivadas, entendiendo estas como un valor numérico que representa cómo dicha función cambia al variar el valor de x . Durante esta explicación, se ha desarrollado el cálculo de derivadas, a partir de sus reglas básicas (polinomios, lineal, coseno, entre otras). Se ha revisado también el cálculo de derivadas en sumas, productos, fracciones y funciones.

Por último, en este primer bloque se han trabajado los conceptos de optimización y de primitivas e integral definida.

En el segundo bloque del capítulo, se ha analizado el concepto de cálculo multivariable, a través de las derivadas parciales, el gradiente y las derivadas direccionales, las segundas derivadas parciales y la matriz hessiana.

Más adelante, se han repasado los conceptos relacionados con la optimización: los puntos de máximo y mínimo (locales y globales) de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, que pueden definirse como las funciones de una variable real, y los puntos críticos de una función, que son los puntos tales que el vector gradiente es igual al vector de ceros.

En el último bloque, se ha realizado un acercamiento a los algoritmos de *gradient descent* y *forward & back propagation*. Este tipo de cálculo más avanzado es utilizado para el entrenamiento de las redes neuronales en inteligencia artificial, a través del aprendizaje multicapa.

Por último, de forma práctica, se ha estudiado como manejar estos conceptos a través de la herramienta TensorFlow, que dispone de una biblioteca de recursos al respecto.



Capítulo 4

Probabilidad y estadística

Introducción

La probabilidad y estadística se basa en el estudio del azar dentro de la realidad, en base a una información previa. En el caso de la inteligencia artificial, se estudian las probabilidades para ayudar a la máquina en el proceso de aprendizaje.

Durante este capítulo se realizará un estudio de los diferentes conceptos de probabilidad y estadística, un acercamiento a la estimación de parámetros, así como a los procesos estocásticos, como son las matrices estocásticas, las cadenas de Markov y los procesos gaussianos.

En último lugar, se realizará, como se ha hecho anteriormente, una aplicación práctica de estos conceptos en la herramienta TensorFlow.

Objetivos

- Conocer los conceptos de probabilidad y estadística y su aplicación en Inteligencia Artificial.
- Desarrollar las operaciones de probabilidad necesarias para establecer predicciones de futuro.
- Aplicar la biblioteca de recursos disponibles en la herramienta TensorFlow sobre algoritmos probabilísticos.

La probabilidad y la estadística se pueden considerar el pilar matemático más importante de la inteligencia artificial y, sobre todo, del aprendizaje automático.

En los problemas de reconocimiento de patrones, los datos de entrenamiento se pueden ver como realizaciones de variables aleatorias bajo una distribución desconocida con el objetivo de tomar algunas decisiones o identificar la distribución que genera los datos. En todos los problemas relacionados con variables que se miden a través de sensores u observaciones, es importante considerar el ruido que estas mediciones conllevan, por lo que es necesario un planteamiento probabilístico de los datos.

Además, en muchos casos es necesario considerar también la aleatoriedad en las respuestas o decisiones que los sistemas automáticos tienen que tomar. Un bot, o mejor dicho, un chatbot es un sistema capaz de conversar de forma similar a un ser humano.



Enlace de interés

Breve artículo de Sarah Mitroff sobre *chatbots* con algunos ejemplos y enlaces a productos en los que se utilizan. Mitroff, S. (5 de mayo, 2016). What is a bot? Here's everything you need to know. *CNET*.

<https://www.cnet.com/how-to/what-is-a-bot>

Estos *bots* no pueden contestar de forma determinista a las conversaciones, sino que se necesita un cierto grado de creatividad y originalidad en ellas. Una de las formas más sencillas y simples de introducir algo que se parezca a la originalidad humana en las máquinas es utilizar la aleatoriedad, es decir, algún tipo de respuesta probabilística.

Los conceptos probabilísticos son también necesarios para formalizar matemáticamente algunos problemas de la inteligencia artificial y para encontrar soluciones óptimas. En particular, todos los métodos estadísticos pueden ser muy útiles para implementar sistemas de inteligencia artificial.

A continuación veremos algunos de los conceptos básicos de la probabilidad y de la estadística para ser capaces de entender y aplicar modelos y técnicas en la inteligencia artificial.

4.1. Introducción a la probabilidad

En este apartado introduciremos los principales conceptos de la probabilidad. El concepto matemático de probabilidad se basa en el concepto de medida y de espacio de medida (de hecho, es un caso particular). Presentaremos los conceptos de probabilidad, variables aleatorias, etc. de forma intuitiva y sin tener que utilizar los conceptos abstractos de medida. De este modo, la exposición será más sencilla, aunque a veces algo imprecisa.



Ejemplo

Consideremos el lanzamiento de una moneda equilibrada. Es adecuado asumir que, al lanzar muchas veces la moneda, la mitad de las veces saldrá cara y la otra mitad cruz. Además, si tuviésemos que apostar entre cara y cruz, nos daría igual qué opción elegir (sabiendo que la moneda es equilibrada).

Imaginemos ahora que tenemos que asignar un valor de certidumbre antes de lanzar la moneda para comunicar nuestra confianza en el evento "Saldrá cara". Supongamos que este valor tenga que estar entre 0 y 1, siendo un valor cercano a 0 sinónimo de baja probabilidad de suceso y un valor cercano a 1 sinónimo de alta probabilidad. Es natural asignar el número 0,5 a ese valor, dado que no podemos decir si "Saldrá cara" es más o menos probable.

En matemáticas, este valor numérico se llama probabilidad y se suele indicar con la letra P (más adelante veremos que la notación puede variar).

Así pues, en el ejemplo de la moneda equilibrada, sabemos que:

$$P(\text{cruz})=0,5$$

$$P(\text{cara})=0,5$$

Consideremos ahora el clásico ejemplo de la urna y las bolas de dos colores. Imaginemos que hay una urna llena de 100 bolas. Sabemos con seguridad que 30 son blancas y 70 negras, y extraemos una bola al azar (las bolas están bien mezcladas). Es razonable decir que la probabilidad de que la bola extraída sea blanca es 0,3.

Espacio muestral y probabilidad

Ahora intentaremos definir algo más formalmente qué es la probabilidad. Empezaremos definiendo el espacio muestral como el conjunto de todos los posibles resultados de un experimento aleatorio. El espacio muestral se suele indicar con la letra S o la letra griega Ω . En el caso del lanzamiento de una moneda, el espacio muestral es $\Omega = \{\text{cara, cruz}\}$; en el caso del lanzamiento de un dado con seis caras, $\Omega = \{1, 2, 3, 4, 5, 6\}$ y en el caso de la urna del ejemplo, $\Omega = \{\text{blanca, negra}\}$. Cada elemento del espacio muestral se llama suceso elemental, y cada subconjunto del espacio muestral se llama suceso o evento.



Ejemplo

Sea $\Omega = \{1, 2, 3, 4, 5, 6\}$ el espacio muestral asociado al lanzamiento de un dado. Consideremos el evento $\Omega \ni E = \{\text{Ha salido un número impar}\} = \{1, 3, 5\}$, donde E es un evento del espacio muestral compuesto por los tres sucesos elementales: 1, 3 y 5.



Ejemplo

En un ejemplo algo más complejo, consideremos el lanzamiento de dos dados de seis caras. En este caso, el espacio muestral es igual a todas las posibles parejas (i, j) , $1 \leq i, j \leq 6$:

$$\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (2, 1), (2, 2), \dots, (3, 1), \dots, (6, 1), \dots, (6, 6)\}$$

Consideremos ahora el evento $E = \{\text{Salen dos números iguales}\}$. Este evento está compuesto por los seis sucesos elementales:

$$E = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

Los eventos de un espacio muestral son subconjuntos, por lo que podemos utilizar las operaciones entre conjuntos para generar otros eventos.

Así pues, si tenemos dos eventos $E, F \subseteq \Omega$, podemos definir el evento unión $E \cup F$, que incluye todos los sucesos elementales que pertenecen a E o a F ; el evento intersección $E \cap F$, que incluye los sucesos que pertenecen a ambos conjuntos; y el evento complementario F^c (o $\neg F$ o \bar{F}), que está formado por todos los eventos que no pertenecen al conjunto inicial F . Además, siempre existe el conjunto vacío \emptyset , que no contiene ningún suceso elemental. Si dos eventos E_1 y E_2 tienen intersección vacía ($E_1 \cap E_2 = \emptyset$), entonces se dice que son dos eventos mutuamente excluyentes. Los eventos de un conjunto son mutuamente excluyentes si cada pareja de eventos en el conjunto es mutuamente excluyente. Es decir E_1, E_2, \dots, E_n son mutuamente excluyentes solo si $E_i \cap E_j = \emptyset$ para cada $i \neq j$, $1 \leq i, j \leq n$.

Una probabilidad P definida sobre el espacio muestral Ω es una función definida sobre los eventos (conjuntos de Ω) con las siguientes propiedades:

- $0 \leq P(E) \leq 1$ para cada evento $E \subseteq \Omega$.
- $P(\Omega) = 1$.
- Para cada pareja de eventos $E_1, E_2 \subseteq \Omega$ tales que $E_1 \cap E_2 = \emptyset$, resulta que $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.

A partir de las propiedades fundamentales (o axiomas) de la probabilidad, se demuestran otras fórmulas útiles:

- $P(E^c) = 1 - P(E)$.
- Si $E_1 \subseteq E_2$, así que $P(E_1) \leq P(E_2)$.
- $P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$
- $P(\emptyset) = 0$.

Observamos ahora que la teoría de la probabilidad es el conjunto de teoremas, resultados y fórmulas que nos permiten calcular las probabilidades de los eventos (y otras cantidades) a partir de las probabilidades de los sucesos elementales. De todos modos, la teoría de la probabilidad no nos permite asignar la probabilidad de los sucesos elementales. Estas probabilidades básicas de cada ejemplo específico son cantidades subjetivas o calculadas a partir de algunas consideraciones físicas o lógicas.

La pareja formada por un espacio muestral y una probabilidad sobre este se llama **espacio de probabilidad**.



Ejemplo

Volvamos al ejemplo del lanzamiento de un dado con seis caras y calculemos la probabilidad del evento $E = \{\text{Ha salido un número impar}\}$. El evento es la unión de los 3 sucesos elementales $\{1\}, \{3\}, \{5\}$. Si asumimos que cada suceso elemental tiene la misma probabilidad, obtenemos que:

$$P(1)=P(2)=\dots=P(6)=\frac{1}{6}$$

Dado que los sucesos elementales son eventos disjuntos entre sí, podemos calcular la probabilidad del evento E :

$$P(E)=P(\{1\} \cup \{3\} \cup \{5\})=P(\{1\})+P(\{3\})+P(\{5\})=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{3}{6}=\frac{1}{2}$$

4.1.1. Independencia de sucesos

El concepto de independencia es uno de los conceptos más importantes en la teoría de la probabilidad, pues nos permite calcular multitud de probabilidades de eventos complejos y es fundamental en el desarrollo de la teoría estadística. En este apartado empezamos a definir el concepto de **eventos independientes**.

Se dice que dos eventos E y F son independientes si:

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

En palabras, la probabilidad de la intersección de dos eventos independientes es igual al producto de las probabilidades de los dos eventos. En la práctica, la independencia de dos eventos se suele asumir a partir de algunas reflexiones de tipo lógico.



Ejemplo

Supongamos que lanzamos dos veces una moneda perfectamente equilibrada y anotamos en secuencia los resultados de los dos lanzamientos. El espacio muestral es entonces:

$$\Omega = \{("cara", "cara"), ("cara", "cruz"), ("cruz", "cara"), ("cruz", "cruz")\}$$

Es sensato decir que los 4 sucesos elementales del espacio Ω tienen todos las mismas probabilidades, igual a $\frac{1}{4}$. Es fácil ver ahora que los eventos $E_1 = \text{cara en el 1.º lanzamiento}$ y $E_2 = \text{cara en el 2.º lanzamiento}$ son dos eventos independientes en este espacio de probabilidad:

$$P(E_1) = P(\{("cara", "cara"), ("cara", "cruz")\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

$$P(E_2) = P(\{("cara", "cara"), ("cara", "cruz")\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Y además:

$$P(E_1 \cap E_2) = P("cara", "cara") = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(E_1)P(E_2)$$

Así pues, los dos eventos son independientes.

También podemos ver este experimento aleatorio como la realización de dos experimentos aleatorios con espacios muestrales $\Omega_1 = \Omega_2 = \{\text{cara, cruz}\}$ y probabilidades de los sucesos elementales iguales a $\frac{1}{2}$, es decir, el lanzamiento de una moneda equilibrada. Entonces podemos asumir que los dos lanzamientos son dos eventos físicos sin ninguna relación entre ellos (es decir, el suceso del primer lanzamiento no influye en el segundo), de modo que los eventos E_1 y E_2 son independientes entre sí y podemos calcular directamente:

$$P("cara", "cara") = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

El concepto de independencia formaliza entonces lo que intuitivamente entendemos al decir que la realización del evento E no influye sobre la realización del evento F . Es importante recordar que la independencia de dos eventos es una asunción en el problema específico y precisamente en el espacio de probabilidad que hemos definido.

4.1.2. Probabilidad condicionada y teorema de Bayes

Otro concepto clave de este ámbito es la probabilidad condicionada, que representa la probabilidad de un evento cuando tenemos la información de que otro evento dado se ha verificado. Formalmente, la probabilidad condicionada de E dado F se define de la siguiente forma:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Esta probabilidad es bien definida para cada evento F de probabilidad distinta de 0 ($P(F) \neq 0$) y define otra probabilidad sobre el mismo espacio muestral: $Q(E) = P(E|F)$.

El teorema de Bayes enlaza las dos probabilidades condicionadas $P(E|F)$ y $P(F|E)$:

$$P(E|F) = \frac{P(E|F)P(F)}{P(F)}$$

A partir del teorema de Bayes podemos demostrar también la fórmula que se llama regla de la cadena:

$$P(E \cap F) = P(E|F)P(F)$$



Ejemplo

Consideremos el lanzamiento de dos dados. Como hemos visto, en este caso, el espacio muestral coincide con el conjunto de las 36 parejas de valores:

$$\Omega = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (2, 6), \dots, (6, 1), \dots, (6, 6)\} = \{(i, j) : 1 \leq i, j \leq 6\}$$

Consideremos los dados perfectamente equilibrados, de modo que para cada pareja de valores posibles $(i, j) \in \Omega$:

$$P(i, j) = \frac{1}{36}$$

Consideremos ahora los eventos $E = \{(i, j) : i = j\}$ y $F = \{i \text{ es par y } j \text{ es par}\}$ y calculemos $P(E|F)$. Intuitivamente, si sabemos que el evento F se ha realizado, entonces sabemos que i y j son ambos números pares. Así pues, enumerando los sucesos elementales en F , obtenemos:

$$F = \{(2, 2), (2, 4), (2, 6), (4, 2), (4, 4), (4, 6), (6, 2), (6, 4), (6, 6)\}$$

Dado que 3 de los 9 eventos de F pertenecen a E , obtenemos que la probabilidad condicionada es igual a $P(E|F) = \frac{1}{3}$. Utilizando directamente la definición para calcular $P(E|F)$, tenemos que calcular el evento intersección:

$$E \cap F = \{(2, 2), (4, 4), (6, 6)\}$$

Y dado que $P(E \cap F) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{12}$ y $P(F) = \frac{9}{36} = \frac{1}{4}$, obtenemos que:

$$P(E|F) = \frac{\frac{1}{12}}{\frac{1}{4}} = \frac{1}{12} \cdot 4 = \frac{1}{3}$$



Ejemplo

Si dos eventos son independientes entre sí, entonces a partir de la definición de probabilidad condicionada obtenemos que:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)P(F)}{P(F)} = P(E)$$

En palabras, la probabilidad condicionada es igual a la probabilidad original si los dos eventos son independientes.

Espacios muestrales infinitos y continuos

Para los espacios de probabilidades con un número infinito de sucesos elementales (o continuos), tenemos que modificar el axioma de la unión en la probabilidad de la siguiente forma:

- Si E_1, E_2, \dots son sucesos mutuamente excluyentes, entonces $P(E_1 \cup E_2 \cup E_3 \cup \dots) = \sum P(E_i) = E_1 + E_2 + E_3 + \dots$

Para los espacios muestrales continuos, la teoría es más compleja y se necesitan conocimientos de teoría de la medida para definir los espacios de probabilidad de forma correcta. Intuitivamente, el problema es que en los espacios muestrales continuos no se puede definir la probabilidad para todos los posibles subconjuntos, lo cual sí es posible para los espacios finitos y numerables. Por esta razón, es necesario definir una clase de subconjunto (los eventos o, mejor dicho, los conjuntos medibles) sobre los cuales es posible definir una probabilidad. Hay distintas formas de definir estos conjuntos medibles, una de las cuales (la más utilizada) es a partir de los intervalos $(a,b) \subset \mathbb{R}$. En la práctica no se suele encontrar ningún problema, sobre todo si los eventos que se utilizan son todos construidos a partir de intervalos (abiertos o cerrados).



Ejemplo

Consideremos la función positiva $g(x) = x^2$ en el intervalo $[0,1]$. Consideremos también el cuadrado Q con vértices $(0,0), (0,1), (1,1), (1,0)$ y área igual a 1. Observamos que la gráfica de la función g está incluida en el cuadrado Q . Imaginemos ahora que elegimos un punto al azar en el cuadrado Q . ¿Cuál es la probabilidad de que el punto esté por debajo de la gráfica de g ?

El espacio muestral es, en este caso, el conjunto de puntos en el cuadrado:

$$\Omega = \{(x,y) : 0 \leq x, y \leq 1\}$$

Dado un evento $A \subseteq \Omega$, podemos definir la probabilidad igual al área de este evento, es decir, $P(A) = \text{Área}(A)$.

Así pues, la probabilidad de que un punto esté por debajo de la gráfica de una función es igual a la integral de la función:

$$P(y \leq g(x)) = \int g(x) dx$$



Ejemplo

$$\text{Y en nuestro ejemplo: } P(y \leq x^2) = \int x^2 dx = \left(\frac{x^3}{3} \right)_0^1 = \frac{1}{3}$$

4.1.3. Variables aleatorias discretas y continuas

Una variable aleatoria es una función real X definida sobre un espacio muestral Ω , $X : \Omega \rightarrow \mathbb{R}$. Para cada realización de un suceso elemental $\omega \in \Omega$, la variable aleatoria asigna un valor real $X(\omega) \in \mathbb{R}$. Si una probabilidad es definida sobre el espacio muestral, entonces es posible inducir una probabilidad sobre la imagen $X(\Omega) \subseteq \mathbb{R}$ a través de la siguiente fórmula:

$$P_x(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega \text{ tal que } X(\omega) \in A\})$$

Donde $X^{-1}(A)$ es la imagen inversa de A a través de la función X .

Las variables aleatorias son el objeto fundamental de la probabilidad y, sobre todo, de la teoría estadística. El espacio muestral Ω representa, en la práctica, el espacio de los posibles resultados de un experimento que, en general, no podemos observar directamente. Lo que podemos observar es una medición que mide un valor numérico para cada realización del experimento aleatorio, es decir, una variable aleatoria.

Variables aleatorias discretas

Si el conjunto $X(\Omega)$ es finito o numerable (es decir, infinito pero no continuo), entonces se dice que la variable aleatoria es una variable aleatoria discreta. En este caso, la probabilidad inducida P_x se caracteriza a partir de las probabilidades de los sucesos elementales. En particular, denotamos con $P_x(x) = f_x(x) = P(X=x) = P(\{\omega \in \Omega : X(\omega) = x\})$ los valores de la **función de densidad** o **función de masa de probabilidad** f_x de la variable aleatoria X . La función de densidad satisface la condición de normalización $\sum f_x(x) = 1$, y la **función de distribución** de la variable aleatoria se define así:

$$F_x(x) = P(X \leq x) = \sum f_x(t)$$



Ejemplo

Sea $\Omega = \{\text{población de España}\}$ y $X : \Omega \rightarrow \mathbb{R}$ la variable aleatoria que asigna a cada persona el año de nacimiento. Está claro que $X(\Omega)$ es un conjunto finito de valores y X es, por lo tanto, una variable aleatoria discreta (finita).



Ejemplo

Consideremos el siguiente experimento aleatorio: lanzamos una moneda varias veces hasta que se obtiene cara y paramos. Sea X la variable aleatoria que cuenta cuántas veces hemos lanzado la moneda antes de obtener cara. $X(\omega) > 0$, pero en principio no hay ninguna cota superior al número de veces que podemos lanzar la moneda y obtener siempre cruz. El espacio muestral $X(\Omega) = \{1, 2, 3, 4, \dots\}$ es, pues, infinito numerable.

Variables aleatorias continuas

Si el conjunto $X(\Omega)$ es continuo (por ejemplo, $X(\Omega) = (a, b)$ o $X(\Omega) = \mathbb{R}$), entonces la variable aleatoria X es una **variable aleatoria continua**. Definimos de forma similar la función de distribución:

$$F_x(x) = P(X \leq x)$$

Decimos que la variable aleatoria admite una densidad si existe una función no negativa e integrable f_x tal que:

$$P(X \in A) = \int f_x(t) dt$$

Donde $A \subseteq \mathbb{R}$ es un evento (un subconjunto) de los números reales. Así pues, podemos escribir para la función de distribución:

$$F_x(x) = \int f_x(t) dt$$

La función de densidad tiene que satisfacer la condición de normalización:

$$\int f_x(t) dt = P(X \in \mathbb{R}) = 1$$

Una de las consecuencias de la existencia de una densidad para una variable aleatoria es que la función de distribución es una función continua:

$$P(X < x) = \lim_{t \rightarrow x} F_x(t) = F_x(x) = P(X \leq x)$$

Dado que $P(X < x) = P(X \leq x)$, la probabilidad de que la variable aleatoria tome un valor $x \in \mathbb{R}$ es igual a 0:

$$P(X = x) = 0 \text{ para cada } x \in \mathbb{R}$$

Esta propiedad es muy importante e indica que, para una variable aleatoria continua (con densidad), la probabilidad de tomar exactamente un valor es siempre igual a 0. Claramente, lo mismo vale para la probabilidad de tomar un conjunto finito o numerable de valores.

4.1.4. Esperanza y varianza de una variable aleatoria

El valor medio o valor esperado (llamado también esperanza o, simplemente, media) de una variable aleatoria es el valor:

$$E[X] = \begin{cases} \sum x P(X=x) & \text{si } X \text{ es una variable aleatoria discreta} \\ \int x f_x(x) dx & \text{si } X \text{ es una variable aleatoria continua} \end{cases}$$

El valor esperado de una variable aleatoria no siempre está bien definido y no siempre existe. Las condiciones bajo las cuales es posible definir el valor absoluto necesitarían una definición más rigurosa de variable aleatoria y de integración que la que utilizamos aquí.

El valor esperado de una variable es una operación lineal:

$$E[X+Y] = E[X] + E[Y] \text{ para cada } X \text{ e } Y \text{ variable aleatoria.}$$

$$E[aX] = aE[X] \text{ para cada } a \in \mathbb{R} \text{ y } X \text{ variable aleatoria.}$$

Además, si dos variables aleatorias son independientes, entonces el valor esperado del producto es igual al producto de los valores esperados:

$$E[XY] = E[X]E[Y] \text{ para cada pareja de variables aleatorias } X, Y \text{ independientes}$$

De forma similar, podemos definir el valor esperado de las funciones de variables aleatorias:

$$E[g(X)] = \begin{cases} \sum g(x)P(X=x) & \text{si } X \text{ es una variable aleatoria discreta} \\ \int g(x)f_x(x) dx & \text{si } X \text{ es una variable aleatoria continua} \end{cases}$$

En particular, nos interesan las funciones del tipo $g(x) = x^n$. Para $n \geq 1$ números naturales, llamamos el valor $E[X^n]$ el **momento de orden n**.

Otra cantidad de interés es la **varianza** de una variable aleatoria, que se define de la siguiente forma:

$$\text{Var}(X) = E[(X-E[X])^2] = E[X^2] - E[X]^2$$

La varianza existe y se puede calcular solo si existe el segundo momento de la variable, en cuyo caso es siempre positiva, y es 0 solo si la variable aleatoria es constante. Además, valen las siguientes propiedades:

- $\text{Var}(aX) = a^2 \text{Var}(X)$ para cada $a \in \mathbb{R}$.
- $\text{Var}(X+a) = \text{Var}(X)$ para cada $a \in \mathbb{R}$.
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$ para cada X, Y independientes.

La varianza de una variable aleatoria es un valor que indica la dispersión alrededor del valor medio. Los valores pequeños de la varianza indican una variable aleatoria muy concentrada, es decir, que la probabilidad de que la variable tome valores cercanos a su media es muy alta. Los valores grandes de la varianza indican una alta dispersión de la variable aleatoria. Este concepto se formaliza en la desigualdad de Chebyshev:

$$P(|X-E[X]| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}$$



Ejemplo

Sea B una variable aleatoria binaria que toma dos valores $\{-a, a\}$ con la misma probabilidad (es decir, $P(X=-a) = P(X=a) = 0,5$). Calculamos la varianza de B en función del valor de $a \geq 0$.

$$\text{Var}(B) = E[B^2] - E[B]^2 = \frac{\text{Var}(X)}{\epsilon^2} \sum b^2 - \left(\sum \frac{1}{2} b \right)^2 = \frac{1}{2} (a^2 + a^2) - \left(\frac{1}{2} (-a + a) \right)^2 = a^2$$

Cuando el valor de a aumenta y claramente la variable aleatoria está más alejada de su media ($E[B] = 0$) es cuando más aumenta la varianza.

La raíz cuadrada de la varianza se llama **desviación estándar** y se suele denotar con la letra griega σ :

$$\sigma_x = \sqrt{(\text{Var}(X))}$$

Si una variable aleatoria X es tal que $E[X^2] < \infty$ y es posible definir la varianza, entonces es posible estandarizar (o normalizar o tipificar) la variable aleatoria, es decir, definir otra variable Z , que es la versión estandarizada (o normalizada o tipificada) de la variable X :

$$Z = \frac{X - E[X]}{\sigma_x}$$

La variable Z se obtiene trasladando X de tal forma que la media sea 0 ($E[Z] = E[X] - E[X] = 0$) y dividiendo por la desviación estándar, de tal forma que:

$$\text{Var}(Z) = 1$$



Ejemplo

Sea U una variable aleatoria continua uniforme entre 0 y 3, es decir, una variable aleatoria con densidad:

$$f_u(t) = \begin{cases} 0 & \text{para } t < 0 \\ \frac{1}{3} & \text{para } 0 \leq t \leq 3 \\ 0 & \text{para } t > 3 \end{cases}$$

Su media es:

$$E[U] = \int t f_u(t) dt = \int \frac{1}{3} \left(\frac{t^2}{2} \right) \Big|_0^3 = \frac{1}{6} (9 - 0) = \frac{3}{2}$$



Ejemplo

Como esperábamos por intuición, el momento de orden 2 es igual a lo siguiente:

$$E[U^2] = \frac{\int t^2 dt}{3} = \frac{1}{3} \left(\frac{t^3}{3} \right) \Big|_0^3 = \frac{1}{9} (27 - 0) = 3$$

La varianza de U es entonces igual a lo siguiente:

$$\text{Var}[U] = E[U^2] - E[U]^2 = 9 - \frac{9}{4} = \frac{3}{4}$$

La versión normalizada de la variable U es entonces igual a $Z = \frac{U - E[U]}{\sqrt{\text{Var}(U)}} = 2 \frac{U - \frac{3}{2}}{\sqrt{3}}$, que es una variable aleatoria continua y uniforme entre $-\frac{3}{2}$ y $\frac{3}{2}$.

Covarianza y matriz de covarianza

La **covarianza** entre dos variables aleatorias X y Y se define así:

$$\text{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$$

La covarianza es un valor que representa cómo varían juntas dos variables aleatorias. Si la covarianza es positiva, entonces los valores mayores de una de las dos variables se corresponden, muy probablemente, con valores mayores de la otra variable. Por el contrario, si la covarianza es negativa, los valores mayores de una de las variables se corresponden con valores menores de la otra.

Si las variables son independientes, entonces la covarianza entre ellas es 0, debido a las propiedades del valor esperado. Es importante observar que hay variables aleatorias que no son independientes entre sí y tales que su covarianza es 0, es decir, que es condición suficiente pero no necesaria.

Las siguientes son las propiedades más importantes de la covarianza:

- $\text{Cov}(X,a) = 0$ para cada $a \in \mathbb{R}$.
- $\text{Cov}(X,X) = \text{Var}(X)$.
- $\text{Cov}(X,Y) = \text{Cov}(Y,X)$.
- $\text{Cov}(aX,bY) = ab\text{Cov}(X,Y)$.
- $\text{Cov}(X+a,Y) = \text{Cov}(X,Y)$.

La covarianza entre las versiones estándar de las variables aleatorias se llama correlación o coeficiente de correlación, que se define directamente así:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \text{Cov}\left(\frac{X}{\sigma_x}, \frac{Y}{\sigma_y}\right)$$

Consideremos ahora un vector de variables aleatorias $X = (X_1, X_2, \dots, X_n)$ y construyamos la siguiente matriz:

$$\Sigma = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \vdots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

Σ se llama matriz de covarianza y es una matriz de dimensión $n \times n$ simétrica semidefinida positiva y cuya diagonal es igual a lo siguiente:

$$\text{diag}(\Sigma) = (\text{Var}(X_1), \dots, \text{Var}(X_n))$$

Si, en vez de la covarianza, consideramos el coeficiente de correlación, entonces obtenemos la matriz de correlación.

4.1.5. Distribuciones discretas y continuas

En este apartado veremos algunos ejemplos de distribuciones de variables aleatorias. La muestra no es completa y solo se pretende presentar algunos ejemplos.

La notación usual para indicar que una variable aleatoria sigue una distribución es $X \sim \text{Distribución}$. Por ejemplo, si X es una variable aleatoria uniforme entre los puntos -1 y 1 , entonces escribiremos $X \sim \text{Unif}([-1, +1])$.

Distribuciones discretas

La distribución discreta más básica es seguramente la distribución de Bernulli. Se dice que una variable aleatoria sigue la distribución de Bernulli de parámetro $p \in [0, 1]$, y escribiremos $X \sim \text{Bernulli}(p)$ si X toma valores en un conjunto binario (se suele asumir $X \in \{0, 1\}$) y:

$$P(X=0) = q = 1-p \quad P(X=1) = p$$

El valor esperado y la varianza de una variable Bernulli se obtienen con las siguientes fórmulas:

$$E[X] = p \quad \text{Var}(X) = p(1-p) = pq$$

Si ahora consideramos n distintas variables aleatorias independientes y cada una con distribución Bernulli de parámetro p , entonces la distribución de la suma $X = X_1 + X_2 + \dots + X_n$ de las variables aleatorias Bernulli se llama distribución binomial y se denota con $X \sim \text{B}(n, p)$. Su distribución de masa (o densidad) se obtiene a través de cálculos de combinatoria y es igual a lo siguiente:

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

La distribución binomial modeliza la probabilidad de éxito en k pruebas de las n totales en una secuencia de experimentos aleatorios binarios.



Ejemplo

Supongamos que lanzamos 10 veces una moneda perfectamente equilibrada. ¿Cuál es la probabilidad de obtener exactamente 4 cruces?

Para contestar, tenemos que observar que la variable aleatoria N números de sucesos sigue una distribución $B(10; 0,5)$. Así pues:

$$P(N=4) = \binom{10}{4} (0,5^4)(0,5^6) = 0,205\dots$$

El valor medio y la varianza de una variable aleatoria $X \sim B(n,p)$ son:

$$E[X] = np \quad \text{Var}(X) = np(1-p)$$

Supongamos ahora que efectuamos un número de pruebas muy alto (en el límite $n \rightarrow +\infty$) y consideramos probabilidades de sucesos $p \rightarrow 0$ tales que:

$$\lim np = \lambda$$

La distribución límite de la distribución binomial con estos parámetros se llama distribución de Poisson y se denota con $X \sim \text{Pois}(\lambda)$. La función de probabilidad es:

$$P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad k=0,1,2,3,\dots$$

La distribución de Poisson se utiliza para modelizar eventos raros, es decir, eventos cuya probabilidad de éxito es muy baja. Además, se utiliza la distribución de Poisson para describir el número de fenómenos (de probabilidad constante) que ocurren en un intervalo de tiempo fijo.

El valor medio y la varianza de una variable aleatoria de Poisson son:

$$E[X] = \lambda \quad \text{Var}(X) = \lambda$$



Ejemplo

Supongamos que en la producción de un componente electrónico específico sabemos que el 1% de las piezas producidas son defectuosas. ¿Cuál es la probabilidad de que 10 de cada 500 piezas producidas sean defectuosas?

Modelizamos el problema a través de la distribución de Poisson. En particular, pongamos X = número de piezas defectuosas de las 500 $\sim \text{Pois}(\lambda)$. El valor de λ se obtiene utilizando el hecho de que $\lambda = E[X] =$ valor promedio de piezas defectuosas sobre 500 = 1% de 500 = 5.

Así pues, la probabilidad de que 10 piezas sean defectuosas es:

$$P(X=10) = \frac{e^{-5} 5^{10}}{10!} \approx 0,018$$

Otra distribución fundamental es la distribución uniforme sobre un número finito de valores. Para simplificar la notación, asumimos que estos valores son los primeros n números enteros positivos $\{1, 2, 3, 4, \dots, n\}$. Una variable aleatoria sigue la distribución uniforme sobre n valores si $X: \Omega \rightarrow \{1, 2, 3, \dots, n\}$ y su función de probabilidad es:

$$P(X=k) = \frac{1}{n}$$

Escribimos $X \sim \text{Unif}(n)$ o $X \sim \text{Unif}(\{1, \dots, n\})$ si queremos explicitar el conjunto de valores.

Distribuciones continuas

Entre las distribuciones continuas, la más importante es seguramente la distribución normal o distribución gaussiana.

Se dice que una variable aleatoria continua es gaussiana o que sigue la distribución normal o gaussiana de parámetros μ y σ^2 (o desviación estándar σ) y escribimos $X \sim N(\mu, \sigma^2)$ si toma valores en todos los números reales y su densidad es:

$$f_x(x) = f_{N(\mu, \sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

El valor medio y la varianza son iguales a los respectivos parámetros $E[X] = \mu$ y $\text{Var}(X) = \sigma^2$. La versión estandarizada de una variable aleatoria es una variable aleatoria normal estandarizada y su densidad es:

$$f_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Las variables aleatorias normales son empleadas con mucha frecuencia en estadística y en aprendizaje automático. Sus propiedades hacen que los cálculos asociados sean muy sencillos y que existan fórmulas cerradas. Además, muchos eventos reales siguen una distribución que se puede modelizar con una distribución normal.

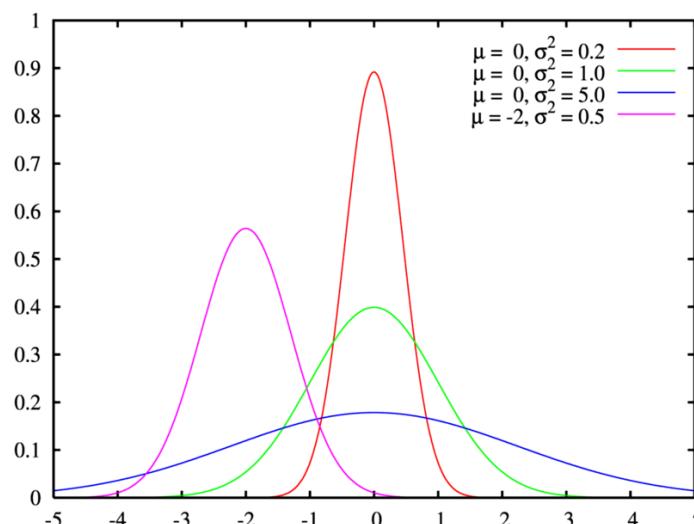


Figura 73. Gráficas de las funciones de densidad gaussianas para varios valores de los parámetros. Por D.328 bajo licencia CC BY-SA 3.0. Fuente: https://commons.wikimedia.org/wiki/File:Normal_distribution_pdf.png



Ejemplo

Supongamos que se quiere modelizar la longitud de una pieza producida en un proceso industrial. Sabemos que las piezas producidas tienen una longitud media de 5 cm y que el error medio de las piezas se mide en $\pm 0,01$ cm. Así pues, podemos modelizar la distribución de la longitud de la pieza con una variable aleatoria $L \sim N(5; 0,001)$.

Los valores de las probabilidades de los eventos de una variable aleatoria gaussiana $X \sim N(\mu, \sigma^2)$ se calculan usando una integral, por ejemplo:

$$P(X \in (a, b)) = \int f_{N(\mu, \sigma^2)}(x) dx$$

En general, estas integrales, aparte de algunos casos particulares, no se pueden resolver con fórmulas cerradas, sino que es necesario utilizar técnicas de aproximación numérica o tablas de valores.

La función de densidad gaussiana es simétrica alrededor de la media, de modo que:

$$P(X < \mu) = \int f_{N(\mu, \sigma^2)}(x) dx = \frac{1}{2} \int f_{N(\mu, \sigma^2)}(x) dx = P(X > \mu)$$

La distribución gaussiana se puede extender a vectores de variables aleatorias. Un vector aleatorio $X = (X_1, \dots, X_n)$ sigue una distribución gaussiana multivariante si la densidad conjunta es igual a lo siguiente:

$$f_X(X_1, \dots, X_n) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2} (x - \mu)^t \Sigma^{-1} (x - \mu)\right)$$

Donde el vector $\mu \in \mathbb{R}^n$ es el vector de los valores medios y Σ es la matriz de covarianza. Las distribuciones marginales de los componentes son gaussianas, sus parámetros son las medias marginales $E[X_i] = \mu_i$ y la varianza $\text{Var}(\Sigma) = \text{diag}(\Sigma)$. Es importante recordar que si un vector aleatorio es gaussiano, entonces sus componentes siguen distribuciones gaussianas. No es verdad la implicación opuesta, es decir, que un vector aleatorio puede tener componentes normales y no seguir una distribución gaussiana multivariante.

Para un vector gaussiano multivariante, es muy sencillo comprobar si sus componentes son independientes. Es suficiente comprobar que las covarianzas sean igual a 0. Es decir, si $(X, Y) \sim N((\mu_x, \mu_y), \Sigma)$ es un vector gaussiano multivariante, entonces X e Y son independientes solo si $\text{Cov}(X, Y) = \Sigma_{1,2} = \Sigma_{2,1} = 0$, o sea, solo si la matriz de covarianza es:

$$\Sigma = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$$

La distribución gaussiana es simétrica y, a veces, se necesitan distribuciones para variables aleatorias no simétricas. Una de las posibilidades es la distribución gamma, cuya función de densidad es:

$$f_{\text{gam}}_{(k, \theta)}(x) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-\frac{x}{\theta}}$$

Donde $\Gamma(t)$ es la función gamma y los parámetros son k (la forma, shape en inglés) y θ (escala). Si $X \sim \text{Gamma}(k, \theta)$, entonces:

$$E[X] = k\theta \quad \text{Var}(X) = k\theta^2$$

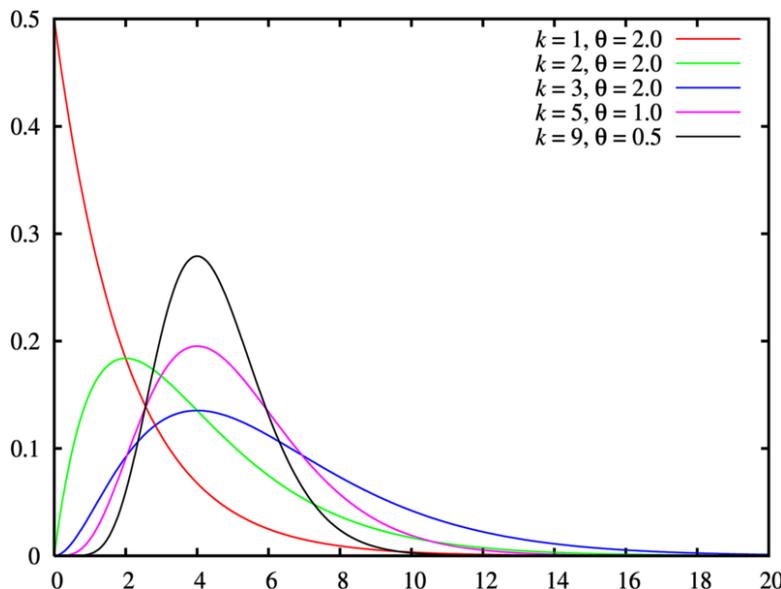


Figura 74. Gráficas de la función de densidad de la distribución gamma para varios valores de los parámetros. Por Auto-pilot bajo licencia CC BY-SA 3.0. Fuente: https://en.wikipedia.org/wiki/Gamma_distribution#/media/File:Gamma_distribution_pdf.svg

Un caso particular de la distribución gamma se da cuando $k=1$ y obtenemos la distribución exponencial.

También para las variables aleatorias continuas podemos definir la distribución uniforme. En particular, decimos que una variable aleatoria es uniforme entre a y b , y escribimos $X \sim \text{Unif}([a,b])$ si la densidad de X es igual a lo siguiente:

$$f_x(x) = \begin{cases} 0 & x < a \\ \frac{1}{(b-a)} & a \leq x \leq b \\ 0 & x > b \end{cases}$$

Es importante recordar que no es posible definir una distribución uniforme sobre todos los números reales, porque, por definición, la densidad de la distribución uniforme es constante sobre el conjunto de valores y, dado que este conjunto no es acotado, la integral de la densidad no puede ser bien definida (es infinita). Esta simple observación es útil para entender que no tiene sentido decir que elegimos un número al azar con probabilidad uniforme si no decimos los límites mínimo y máximo.

Para los vectores aleatorios, es posible generalizar la distribución uniforme sobre cualquier conjunto. Es decir, $X \in \mathbb{R}^n$ es un vector aleatorio con distribución uniforme sobre $A \subseteq \mathbb{R}^n$, un conjunto acotado, si la función de densidad es igual a lo siguiente:

$$f_x(x) = \begin{cases} \frac{1}{\text{Área}(A)} & x \in A \\ 0 & x \notin A \end{cases}$$

4.2. Estimación de parámetros

En este apartado nos familiarizaremos con algunos conceptos de estadística básica. La estadística es la aplicación de la teoría de la probabilidad al estudio de una población a partir de un pequeño conjunto (una muestra).

Uno de los objetivos es estimar la distribución que han generado algunos datos reales para sucesivamente poder efectuar algunas estimaciones o pronósticos. En particular, en este apartado nos centraremos en la estimación de los parámetros de una distribución.

Un concepto fundamental en estadística es el de **variables independientes e idénticamente distribuidas (i. i. d.)**. Se dice que las variables X_1, \dots, X_n aleatorias de un conjunto son independientes e idénticamente distribuidas si $X_i \sim X_j$, es decir, si las variables tienen la misma distribución ($P(X_i < x) = P(X_j < x)$) y son independientes entre sí.



Ejemplo

Supongamos que lanzamos 100 veces un dado equilibrado. Así pues, podemos indicar X_i con la variable aleatoria relacionada con el i -ésimo lanzamiento. Está claro que las variables x_1, \dots, x_n son independiente e idénticamente distribuidas.

El concepto de variables i. i. d. es fundamental para definir el concepto de muestra. Supongamos que observamos un número finito de realizaciones de una variable aleatoria: $X, x_1, x_2, x_3, \dots, x_n$. Podemos asumir que cada una de las realizaciones es una variable aleatoria i. i. d. con la misma distribución que X .

Se dice que un conjunto de puntos $x_1, x_2, x_3, \dots, x_n$ es una muestra de X . El objetivo de las técnicas estadísticas es extraer conocimiento, es decir, información, sobre la distribución de X (que suponemos ahora desconocida) a partir de la muestra. A partir de una muestra es posible calcular algunos valores, los más usuales de los cuales son la media muestral y la matriz de covarianza muestral.

La **media muestral** de la muestra $x_1, x_2, x_3, \dots, x_n$ es la media aritmética:

$$\bar{x} = \frac{\sum x_i}{n}$$

La fórmula es válida también para los vectores de variables aleatorias. En este caso, la media muestral es también un vector de la misma dimensión.

En general, es posible definir el momento muestral de orden λ :

$$\mu_k = \frac{1}{n} \sum x_i^k$$

Está claro que el momento muestral de orden 1 es la media muestral.

La **matriz de covarianza muestral** se calcula a partir de una muestra $x_1, x_2, x_3, \dots, x_n$ de vectores, es decir, $x_i \in \mathbb{R}^d$.

La matriz de covarianza muestral se calcula, pues, usando la siguiente fórmula:

$$\hat{\Sigma} = \frac{1}{n-1} \sum (x_i - \bar{x})(x_i - \bar{x})^t$$

El denominador es $(n-1)$ en vez de n por algunas razones teóricas que no abordaremos aquí. Merece solo la pena recordar que, por ejemplo, para la varianza existen las dos formas de calcular la varianza muestral:

$$s_n^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

4.2.1. Límites de variables aleatorias

En este apartado veremos dos resultados fundamentales de la estadística y la probabilidad. Ambos resultados estudian cómo la media muestral converge al valor medio de una variable aleatoria.

El primer resultado es la ley de los grandes números, que intuitivamente dice que la media muestral converge, cuando el tamaño de la muestra crece, al valor medio de la variable aleatoria. En particular, si X_1, X_2, \dots es una sucesión de variables aleatorias independientes con el mismo valor esperado μ y la misma varianza σ^2 (cuando las variables son i. i. d., satisfacen estas condiciones), si denotamos $\bar{X}_n = (\sum X_i)/n$ con el promedio de las primeras n variables, entonces:

$$\lim_{n \rightarrow \infty} P(\bar{X}_n - \mu | < \delta) = 1 \text{ para cada } \delta > 0$$

Es decir, \bar{X}_n converge en probabilidad a μ .

El segundo resultado, y probablemente el resultado fundamental de la estadística, es el teorema del límite central. Consideremos la versión estandarizada de la suma de variables aleatorias i. i. d.:

$$S_n = X_1 + X_2 + \dots + X_n$$

$$Z_n = \frac{S_n - n\mu}{\sigma \sqrt{n}}$$

Donde μ, σ son la media y la desviación estándar de las variables aleatorias X_i . Así pues:

$$\lim_{n \rightarrow \infty} P(Z_n < t) = P((0,1) < t)$$

Es decir, el límite de la suma estandarizada de variables aleatorias i. i. d. se distribuye como una normal estándar. Podemos también expresar el resultado en función de la media muestral X , en particular si \bar{X}_n es la media muestral de las primeras n observaciones, de modo que \bar{X}_n tiene aproximadamente (para n suficientemente grande) una distribución gaussiana

$$N\left(\mu, \frac{\sigma^2}{n}\right).$$

4.2.2. Método de los momentos

El método de los momentos para estimar los parámetros de una distribución se basa en resolver las ecuaciones obtenidas igualando los momentos muestrales con los momentos teóricos. En algunos de los ejemplos precedentes ya hemos utilizado de alguna forma esta técnica para calcular los parámetros de nuestros modelos.

Si el modelo de probabilidad se especifica a partir de k parámetros, entonces el método de los momentos equivale a resolver las siguientes ecuaciones en los parámetros:

$$\left\{ \begin{array}{l} \mu_1 = E[X] \\ \mu_2 = E[X^2] \\ \vdots \\ \mu_k = E[X^k] \end{array} \right.$$

Donde μ_i es el momento muestral de orden i .



Ejemplo

Supongamos que obtenemos la muestra 1;3;2;2;5;3;7;1;5;4 y supongamos que nuestro modelo es una variable aleatoria uniforme $X \sim \text{Unif}[0,2]$. ¿Cuál es el estimador del método de los momentos de θ ?

Para contestar es suficiente calcular la media muestral \bar{x} y el valor medio de la variable $E[X]$, y resolver la ecuación $E[X] = \bar{x}$.

$$\bar{x} = \frac{(1+3+2+2+5+3+7+1+5+4)}{6} = 2,65$$

$$E[X] = \frac{(\theta - 0)}{2} = \frac{\theta}{2}$$

Así pues, tenemos que resolver $\frac{\theta}{2} = 2,65$, que nos proporciona la estimación:

$$\hat{\theta} = 5,3$$



Ejemplo

Sea $Y \sim N(\mu, \sigma^2)$ una variable gaussiana con media y varianza desconocidas, y supongamos que obtenemos una muestra i. i. d. tal que la media muestral es igual a 5 y la varianza muestral es igual a 2. Podemos, entonces, aplicar el método de los momentos para encontrar los estimadores $\hat{\mu}$ y $\hat{\sigma}$. En particular:

$$E[X] = \hat{\mu} = \bar{x} = 5$$

$$E[X^2] = \hat{\sigma}^2 + E[X]^2 = \hat{\sigma}^2 + \bar{x}^2 = s_n^2 + \bar{x}^2$$

Resolviendo el sistema de dos ecuaciones en las incógnitas $\hat{\mu}$, $\hat{\sigma}$ obtenemos:

$$\hat{\mu} = \bar{x} = 5, \hat{\sigma}^2 = s_n^2 = 2$$

4.2.3. Método de la máxima verosimilitud

El método de los momentos es, en general, bastante sencillo de aplicar, pero no tiene buenas propiedades teóricas, por lo que se suele utilizar el **método de máxima verosimilitud** para estimar los parámetros de una distribución.

La **verosimilitud** de un modelo, dada una muestra de observaciones, es:

$$L(\theta | x_1, x_2, \dots, x_n) = f_x(x_1 | \theta) f_x(x_2 | \theta) \dots f_x(x_n | \theta) = \prod f_x(x_i | \theta)$$

Donde $f_x(x | \theta)$ es la función de densidad (o de probabilidad) de la variable aleatoria en función de los parámetros θ .

El estimador de máxima verosimilitud (MLE) $\hat{\theta}_{MLE}$ de los parámetros del modelo se encuentra maximizando la verosimilitud, es decir, encontrando el punto de máximo de la función $L(\theta | x_1, \dots, x_n)$ vista como función de los parámetros θ .

Intuitivamente, la verosimilitud representa la probabilidad de la secuencia de observaciones X_1, \dots, X_n bajo la hipótesis de que las observaciones son i. i. d. con distribución $f_x(x | \theta)$. Así pues, maximizar la verosimilitud es equivalente a encontrar el modelo (en la familia paramétrica de modelos considerados) que consiga la mayor probabilidad de que se realicen las observaciones.

Encontrar el punto de máximo de la verosimilitud es equivalente a encontrar el punto de máximo de la log-verosimilitud, es decir, la función $LL(\theta | x_1, \dots, x_n)$. La ventaja de la log-verosimilitud es que, por las propiedades del logaritmo, el producto en la definición de la verosimilitud se transforma en una suma, de modo que la log-verosimilitud es, en general, mucho más sencilla de derivar (para encontrar los puntos críticos).

$$LL(\theta | x_1, \dots, x_n) = \sum \log(f_x(x_i | \theta))$$



Ejemplo

Supongamos que obtenemos n observaciones x_1, \dots, x_n de una variable gaussiana con media desconocida y varianza igual a 1, es decir, $X \sim N(\mu, 1)$. La log-verosimilitud es entonces igual a lo siguiente:

$$LL(\mu | x_1, \dots, x_n) = \sum \log f_{N(\mu, 1)}(x_i)$$

Donde $\log f_{N(\mu, 1)}(x) = \log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{(x-\mu)^2}{2}$, de modo que:

$$LL(\mu | x_1, \dots, x_n) = \sum \log f_{N(\mu, 1)}(x_i) = n \log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{\sum (x_i - \mu)^2}{2}$$

Para encontrar el punto de máximo (si existe), podemos derivar y encontrar los puntos críticos:

$$\frac{dLL(\mu | x_1, \dots, x_n)}{d\mu} = 0 - \sum (x_i - \mu)$$

Para encontrar los puntos críticos, buscamos ahora donde la derivada es 0:

$$\sum (x_i - \mu) = 0$$

$$\sum x_i = n\mu$$

$$\mu = \frac{\sum x_i}{n} = \bar{x}$$

El único punto crítico es $\mu = \bar{x}$, que fácilmente vemos que es un punto de máximo (la derivada segunda es siempre negativa). Observamos que el estimador de máxima verosimilitud es igual al estimador por el método de los momentos (aunque no siempre es así).



Ejemplo

Sea $X \sim \text{Bernoulli}(p)$ una variable aleatoria Bernoulli con parámetro desconocido y supongamos que obtenemos una muestra i. i. d. a partir de X .

Para calcular el estimador de máxima verosimilitud \hat{P}_{MLE} , consideraremos la log-verosimilitud:

$$LL(p | x_1, \dots, x_n) = \sum \log P(X=x_i) = \sum \log(1-p) + \sum \log(p)$$

Por tanto, si calculamos la derivada:

$$\frac{dLL(p | x_1, \dots, x_n)}{dp} = \frac{n_0}{1-p} - \frac{n_1}{p}$$

Donde n_0, n_1 son el número de puntos en la muestra con valores 0 y 1, respectivamente, así que el estimador MLE se obtiene resolviendo la ecuación asociada para encontrar los puntos críticos:

$$\hat{p}_{MLE} = \frac{n_1}{n}$$

4.3. Procesos estocásticos

Lo que entendemos por proceso estocástico es un sistema compuesto por variables aleatorias que varían con el tiempo. Unos ejemplos de estos procesos pueden ser:

- Emisión de radio
- Electrocardiograma
- El tiempo de espera del próximo autobús.
- El tiempo que tardará un jugador en pasarse una pantalla en concreto de un videojuego.

A los posibles valores que pueden tomar dichas variables los conocemos como **estados** y se pueden clasificar dependiendo del tipo de espacio en el que se muevan:

- Discretos
- Continuos

Dependiendo del espacio mencionado anteriormente en el que se pueden mover las variables X y de cómo sea el conjunto del tiempo T para ese movimiento, tendremos las siguientes combinaciones:

X\T	Discreto	Continuo
Discreto	Cadena	Procesos de Saltos Puros
Continuo	Sucesión de Variables Aleatorias	Proceso Continuo

4.3.1. Matrices estocásticas

Podemos definir como matriz estocástica a aquella matriz cuadrada $M \in \mathbb{R}^{n \times n}$ tal que cumple las siguientes condiciones:

1. Para todo elemento de M, se cumple que $a_{ij} \geq 0$
2. Se cumple que:

n

$$\sum_{j=1}^n a_{ij} = 1 \quad \text{para cada } i \text{ fijo}$$

Un ejemplo de matriz que satisface estos requisitos serían las matrices identidad:

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

Este tipo de matrices son frecuentemente utilizadas para definir las transiciones de las cadenas de Markov.

4.3.2. Cadenas de Markov

Vamos a retomar los procesos estocásticos y, más concretamente, aquellos de estado discreto. Tomando como punto de partida que los valores en cada instante t de una variable X aleatoria se definirían como:

$$\{X_0=x_0, X_1=x_1, \dots, X_{n-1}=x_{n-1}, X_n=x_n\}$$

Tenemos que la **probabilidad de ocupación** de un estado se define por:

$$P(X_n=x_n)$$

De tal forma que llamemos **probabilidad de cambio de estado** o transición a la probabilidad condicionada siguiente:

$$P(X_n=x_n/X_{n-1}=x_{n-1})$$

Cuando un proceso cumple que la probabilidad de cambiar de estado depende solamente del estado actual en el que nos encontremos y no de todas las transiciones anteriores, se conoce como **propiedad de Markov**.

Recordemos ahora la tabla que definimos para los procesos estocásticos: si teníamos un espacio discreto en un tiempo discreto, a esos procesos los llamábamos “cadenas”. Cuando unas cadenas cumplen con la propiedad de Markov, las llamamos **cadenas de Markov**.

Vamos a suponer, como ejemplo, el caso de un alumno que haya traído sus deberes hechos a clase diariamente. Como punto de partida, los estados disponibles serían:

- Los ha hecho → 1
- No los ha hecho → 0

Por lo tanto, vamos a suponer que si no los ha hecho hoy, lo único que puede suceder mañana es que los haga o que siga sin hacerlos. Para ello, si a la probabilidad de que mañana tampoco los haga la llamamos p , entendemos que, para el caso contrario, dicha probabilidad será $1-p$. De forma análoga, tenemos que si hoy los ha hecho, la probabilidad de que mañana también los haga puede ser q , mientras que el caso contrario será $1-q$. De esta forma, podemos definir una matriz de probabilidades de transición tal que:

$$P = \begin{pmatrix} P(X_{n+1}=0/X_n=0) & P(X_{n+1}=1/X_n=0) \\ P(X_{n+1}=0/X_n=1) & P(X_{n+1}=1/X_n=1) \end{pmatrix} = \begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix}$$

La matriz de Markov se define como una matriz cuadrada tal que $P_{n \times n}$, donde n será el número de estados posibles a transicionar. De forma que si tenemos una matriz definida de la siguiente manera:

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{pmatrix}$$

Donde para cada elemento p_{ij} y siendo un estado. Tenemos que:

$$P_{ij} = P(X_n = s_j / X_{n-1} = s_i)$$

Por último, destacar que esto nos facilita conocer las probabilidades de más de un paso, lo cual nos recuerda mucho a los procesos iterativos que ya conocemos para definir una red neuronal. Para una transición de dos pasos, tenemos que:

$$P^2 = P \times P$$

Siendo de forma análoga la transición a 3, 4... k pasos, definidas como P^3, P^4, \dots, P^k , siendo P^k conocida como la matriz de transición de k pasos de la cadena de Markov.

4.3.3. Procesos gaussianos

Los procesos gaussianos son utilizados cuando el problema que intentamos resolver se mueve en cualquier momento del tiempo; es decir, son procesos en tiempo continuo. El uso de los procesos gaussianos surge de la necesidad de no solamente intentar predecir un valor para un modelo concreto, sino de aportar también un intervalo de confianza o incertidumbre, lo que proviene de la distribución gaussiana. Son muy utilizados para sistemas de múltiples salidas.

Podemos decir que un proceso estocástico con una continuidad en el tiempo es gaussiano si y solo si para cada conjunto finito de índices $(t_1, t_2, \dots, t_k) \in T$, tal que:

$$X_{t_1, \dots, t_k} = (X_{t_1}, \dots, X_{t_k})$$

Siendo este el vector de una variable aleatoria gaussiana, podemos formular la propiedad gaussiana:

Es Gaussiana si y solo si para cada conjunto finito de índices $(t_1, t_2, \dots, t_k) \in T$ existen valores reales y positivos de μ tales que se cumpla la siguiente igualdad:

$$E\left[\exp\left(i \cdot \sum_{l=1}^k t_l X_{t_l}\right)\right] = \exp\left(-\frac{1}{2} \sum_{l,j} \sigma_{lj} t_l t_j + \sum_l \mu_l t_l\right)$$

Donde:

- $E()$ es la esperanza matemática.
- i es el número imaginario igual a $\sqrt{-1}$
- σ_{lj} y μ_l son la covarianza y la media del proceso.

4.4. Desarrollo de algoritmos probabilísticos con TensorFlow

Ahora vamos a ver algunos ejemplos de desarrollo de los conceptos que acabamos de analizar. Para ello, haremos uso de Python y su librería NumPy, que, además de ayudarnos con la construcción de vectores y matrices, cuenta con otras funciones matemáticas avanzadas que nos van a venir muy bien para poder estudiar los algoritmos probabilísticos con TensorFlow.

Partimos de la base de que el alumno cuenta con algunos conocimientos sobre Python y la importación de librerías para la ejecución de *scripts*. En cualquier caso, para el caso que vamos a tratar, lo que tendríamos que tener de base es:

```
from pprint import pprint
import matplotlib.pyplot as plt
import numpy as np

import tensorflow.compat.v2 as tf
tf.enable_v2_behavior()

import tensorflow_probability as tfp

plt.rcParams['image.cmap'] = 'viridis'

%matplotlib inline

tfd = tfp.distributions
```

Figura 75. Código Python para importar las librerías necesarias. Fuente de los datos utilizados para obtener la representación gráfica: https://colab.research.google.com/github/tensorflow/probability/blob/master/tensorflow_probability/examples/jupyter_notebooks/A_Tour_of_TensorFlow_Probability consultado en 12/04/2021

Vamos a empezar con el proceso gaussiano. Vamos a definir una función *kernel* con una exponencial cuadrática o exponencial al cuadrado, para un proceso lineal de gaussiano, desde -4 hasta 4 con 250 puntos:

```
kernel = tfp.math.psd_kernels.ExponentiatedQuadratic()
xs = np.linspace(-4., 4., 250).reshape([-1, 1])
gp = tfd.GaussianProcess(kernel, index_points=xs)
print("Forma del Batch:", gp.batch_shape)
print("Forma del evento:", gp.event_shape)

Forma del Batch: ()
Forma del evento: (250,)
```

Figura 76. Código Python para definir el proceso gaussiano. Fuente de los datos utilizados para obtener la representación gráfica: https://colab.research.google.com/github/tensorflow/probability/blob/master/tensorflow_probability/examples/jupyter_notebooks/A_Tour_of_TensorFlow_Probability consultado en 12/04/2021

Ahora veamos cómo nos quedaría gráficamente este proceso en el que se considera una $\mu=0$

```
upper, lower = gp.mean() + [2 * gp.stddev(), -2 * gp.stddev()]
plt.plot(xs, gp.mean())
plt.fill_between(xs[...], 0, upper, lower, color='k', alpha=.1)
for _ in range(5):
    plt.plot(xs, gp.sample(), c='r', alpha=.3)
plt.title(r"Proceso Gaussiano de media previa, intervalos  $2\sigma$ , y muestras")
plt.show()
```

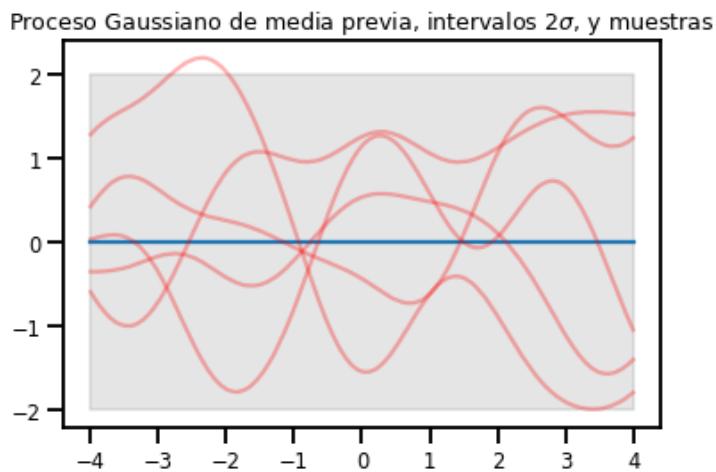


Figura 77. Representación gráfica del proceso gaussiano definido. Fuente de los datos utilizados para obtener la representación gráfica: https://colab.research.google.com/github/tensorflow/probability/blob/master/tensorflow_probability/examples/jupyter_notebooks/A_Tour_of_TensorFlow_Probability consultado en 12/04/2021

Ahora vamos a generar para el mismo *kernel*, un proceso gaussiano en regresión. Obtenemos:

```
# Supongamos que tenemos algunos datos observados
obs_x = [[-3.], [0.], [2.]] # Forma 3x1 (3 1-D Vetores)
obs_y = [3., -2., 2.]       # Forma 3   (3 Escalares)

gprm = tfd.GaussianProcessRegressionModel(kernel, xs, obs_x, obs_y)

upper, lower = gprm.mean() + [2 * gprm.stddev(), -2 * gprm.stddev()]
plt.plot(xs, gprm.mean())
plt.fill_between(xs[...], lower, upper, color='k', alpha=.1)
for _ in range(5):
    plt.plot(xs, gprm.sample(), c='r', alpha=.3)
plt.scatter(obs_x, obs_y, c='k', zorder=3)
plt.title(r"GP media posterior, intervalos  $2\sigma$  intervals, y muestras")
plt.show()
```

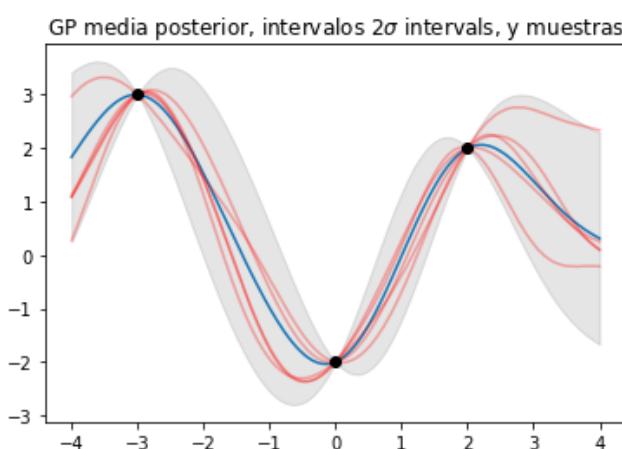


Figura 78. Representación gráfica del proceso gaussiano definido. Fuente de los datos utilizados para obtener la representación gráfica: https://colab.research.google.com/github/tensorflow/probability/blob/master/tensorflow_probability/examples/jupyter_notebooks/A_Tour_of_TensorFlow_Probability consultado en 12/04/2021

Es notable la diferencia cuando usamos la regresión. Vemos también que todas convergen en los puntos definidos en *obs_x* y *obs_y*.

Veamos ahora un ejemplo del algoritmo de Monte Carlo para las cadenas de Markov:

```
# Generamos algunos datos
def f(x, w):
    # Rellenamos x con 1's para añadir sesgo a través de matmul
    x = tf.pad(x, [[1, 0], [0, 0]], constant_values=1)
    linop = tf.linalg.LinearOperatorFullMatrix(w[..., np.newaxis])
    result = linop.matmul(x, adjoint=True)
    return result[..., 0, :]

num_features = 2
num_examples = 50
noise_scale = .5
true_w = np.array([-1., 2., 3.])

xs = np.random.uniform(-1., 1., [num_features, num_examples])
ys = f(xs, true_w) + np.random.normal(0., noise_scale, size=num_examples)

# Visualizar el conjunto de datos
plt.scatter(*xs, c=ys, s=100, linewidths=0)

grid = np.meshgrid(*([np.linspace(-1, 1, 100)] * 2))
xs_grid = np.stack(grid, axis=0)
fs_grid = f(xs_grid.reshape([num_features, -1]), true_w)
fs_grid = np.reshape(fs_grid, [100, 100])
plt.colorbar()
plt.contour(xs_grid[0, ...], xs_grid[1, ...], fs_grid, 20, linewidths=1)
plt.show()
```

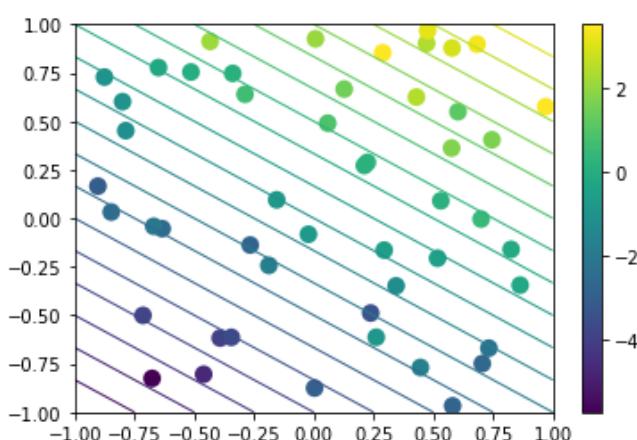


Figura 79. Configuración de datos y visualización para una cadena de Markov. Fuente de los datos utilizados para obtener la representación gráfica: https://colab.research.google.com/github/tensorflow/probability/blob/master/tensorflow_probability/examples/jupyter_notebooks/A_Tour_of_TensorFlow_Probability consultado en 12/04/2021

La idea del método MCMC (algoritmo de Monte Carlo para las cadenas de Markov) es simular una cadena de Markov para una distribución estacionaria. Tengamos en cuenta algunas ideas:

- Al igual que en las cadenas de Markov, cada valor simulado va a depender exclusivamente del valor que le precede.
- Independientemente de los valores que elijamos para empezar, la convergencia estará asegurada.
- Hacen falta un alto número de iteraciones para asegurarse una aproximación estacionaria de la distribución.
- Los primeros valores suelen ser eliminados por no estar cerca del estado estacionario.

Resumen

En Inteligencia Artificial, es necesario no solo que la máquina como tal realice relaciones lógicas con la información suministrada, sino que aprenda a predecir el futuro comportamiento de la realidad. Estas decisiones se basan en las teorías matemáticas de probabilidad y estadística.

Durante el primer bloque se ha examinado el concepto de probabilidad, que está relacionado con el concepto de medida y espacio de medida, aunque, en este caso, se realizó la presentación de los diferentes conceptos de forma intuitiva, para que resultara más sencillo.

De esta manera se han trabajado los conceptos de espacio muestral y probabilidad, además del de suceso elemental y del subconjunto suceso (o evento). Se ha desarrollado también el concepto de independencia de sucesos y de probabilidad condicional y el teorema de Bayes. En otros términos, se ha revisado también el concepto de espacios muestrales infinitos y continuos.

En cuanto a las variables que pueden intervenir en el resultado final, se ha analizado el concepto de variables aleatorias discretas y continuas, como elementos básicos, así como el de esperanza y varianza.

Como última parte de este primer bloque, se ha hablado de los conceptos de distribuciones discretas y continuas.

En un segundo bloque, se ha desarrollado el trabajo con estimaciones de parámetros, para el que es necesario dominar los conceptos de límites de variables aleatorias y dominar los métodos de los momentos, por un lado, y, por otro, el método de la máxima verosimilitud.

En el tercer bloque, se han examinado los procesos estocásticos, a través de las matrices estocásticas, la cadena de Markov y, por último, los procesos gaussianos.

Ya en el último bloque se ha desarrollado la aplicación práctica de estos procesos y conceptos a través de la herramienta de código libre TensorFlow que, a través de su librería de herramientas, ha permitido comprobar cómo puede programarse una máquina para que aprenda de los datos suministrados y prevea el azar, a través de las herramientas estadísticas estudiadas en este capítulo.

Glosario

Autovalor

Valor numérico λ asociado a una matriz por la fórmula $\lambda v = Av$, donde $v \in R^n$ es el autovector correspondiente.

Autovector

Vector $v \in R^n$ correspondiente a un autovalor $\lambda \in R$ tal que $\lambda v = Av$

Codominio

Conjunto de valores que una función puede asumir. En otras palabras, las salidas de la función correspondiente pueden solo ser elementos del codominio. Es uno de los tres elementos que definen una función.

Correlación

Coeficiente numérico que indica la relación entre dos variables aleatorias. Los valores positivos (o negativos) indican que, al aumentar una de las variables, la otra muy probablemente crece (o decrece). Es un valor entre -1 y 1, y se computa como la covarianza entre las versiones normalizadas de las dos variables aleatorias.

Covarianza

Valor que, de forma similar a la correlación, indica la relación entre dos variables. Se computa como $Cov(X,Y) = E[XY] - E[X]E[Y]$.

Definida negativa

Una matriz es definida negativa si $x^t Ax < 0$ para todos los vectores $x \in R^n$ distintos del vector de ceros.

Definida positiva

Una matriz es definida positiva si $x^t Ax > 0$ para todos los vectores $x \in R^n$ distintos del vector de ceros.

Derivada de una función

La derivada de una función en un punto es un valor numérico que indica cómo la función cambia en un entorno. Se define a través de un límite y se puede interpretar geométricamente como el coeficiente de la recta tangente a la gráfica de la función.

Derivada direccional

La derivada direccional es la derivada de una función de más de una variable en una dirección precisa. Se computa con el producto escalar del gradiente por la dirección elegida.

Derivada parcial

La derivada parcial de una función de distintas variables es la derivada respecto a una de las variables fijando las demás.

Desviación estándar

La desviación estándar es una medida de concentración de las variables aleatorias. En particular, se define como la raíz cuadrada de la varianza.

Determinante

El determinante de una matriz cuadrada es un valor numérico esencial en el álgebra lineal. Una propiedad fundamental es que una matriz es invertible solo si su determinante es distinto de 0.

Diagonal de la matriz

La diagonal de una matriz cuadrada $A \in R^n$ es el vector de elementos $a_{1,1}, a_{2,2}, \dots, a_{n,n}$.

Dominio

El dominio de una función es el conjunto de valores de partida de una función.

Espacio de probabilidad

Un espacio de probabilidad es un espacio muestral, es decir, un espacio de eventos, junto con una función de probabilidad definida sobre ellos.

Espacio muestral

Un espacio muestral es un espacio sobre el que se define una probabilidad. Sus subconjuntos son los eventos o sucesos, y sus elementos se llaman sucesos elementales.

Espacio vectorial

Un espacio vectorial es un conjunto de vectores cerrados con respecto a las operaciones de suma y de producto por escalar.

Esperanza

Véase *valor medio*.

Evento

Un evento es un subconjunto de un espacio muestral.

Eventos independientes

Dos eventos en un espacio de probabilidad son independientes si la probabilidad del evento intersección es igual al producto de las probabilidades de los dos eventos.

Eventos mutuamente excluyentes

Dos eventos son mutuamente excluyentes si su intersección es el evento vacío.

Extremos de integración

Los extremos de integración de una integral son los valores $a, b \in \mathbb{R}$ tales que la integral se escribe como $\int f(x)dx$.

Función

Una función es una regla que asocia a cada valor en un conjunto de partida (el dominio) un solo elemento en un conjunto de llegada, llamado codominio. Una función se define por tres elementos: la regla, el dominio y el codominio.

Función continua

Una función de variable real es continua si en cada punto de su dominio el límite en el punto existe y es igual al valor de la función en el mismo punto. La gráfica de una función continua es una línea ininterrumpida.

Función de densidad

La función de densidad de una variable aleatoria continua es una función f positiva y tal que $P(X < t) = \int f(x)dx$. Para las variables aleatorias discretas, la función de densidad es igual a la función de masa de probabilidad.

Función de distribución

La función de distribución de una variable aleatoria es la función $F_x(x) = P(X \leq x)$.

Función de masa de probabilidad

La función de masa de probabilidad, o simplemente función de probabilidad, de una variable aleatoria discreta es la función $f_x(x) = P(X=x)$, es decir, es la probabilidad de los sucesos elementales.

Función monótona creciente

Una función es monótona creciente si $f(x_1) < f(x_2)$ para cada $x_1 < x_2$.

Función monótona decreciente

Una función es monótona decreciente si vale $f(x_1) > f(x_2)$ para $x_1 < x_2$.

Gráfica de función

La gráfica de una función de variables reales es el conjunto de puntos del tipo $(x, f(x))$. La gráfica de una función es una imagen en dos dimensiones para las funciones de una variable y es en tres dimensiones para las funciones de dos variables.

Gradiente de una función

El gradiente es el vector formado por las derivadas parciales de una función de más de una variable.

Imagen de la función

La imagen de una función es el conjunto de puntos en el codominio, que son el resultado de computar la función en algún punto del dominio. Es siempre un subconjunto del codominio.

Imagen inversa

La imagen inversa de un subconjunto de puntos del codominio de una función es $f^{-1}(C)=\{x \in A: f(x) \in C\}$. Es decir, es igual al conjunto de puntos cuya salida por la función termina en un subconjunto C del codominio.

Integral de una función

La integral de una función es una operación que se denota con $\int f(x)dx$ e intuitivamente está relacionada con el área que hay debajo de la gráfica de la función.

Integral indefinida

La integral indefinida es la integral de una función en que el extremo de integración superior es variable. La integral indefinida es, pues, una función del extremo de integración de la integral.

Límite de una función

El límite de una función es un concepto matemático que intuitivamente representa el valor de la función considerada cuando la variable se acerca a un punto.

Mapa

Véase *función*.

Matriz

Una matriz es un objeto matemático que se representa como una tabla de valores numéricos con dos dimensiones: la primera dimensión es el número de filas y la segunda el número de columnas.

Matriz cuadrada

Una matriz cuadrada es una matriz con el mismo número de filas y de columnas, es decir, con dimensiones iguales.

Matriz de covarianza muestral

La matriz de covarianza muestral es una estimación de la matriz de covarianza que se calcula a partir de los datos de la muestra.

Matriz identidad

La matriz identidad es una matriz cuadrada compuesta de ceros en todos los elementos excepto en la diagonal, donde los elementos son unos. La matriz identidad se denota con el símbolo I_n .

Matriz ortogonal

Una matriz es ortogonal si su inversa es igual a la matriz traspuesta, es decir, $AA^t = A^tA = I_n$.

Matriz traspuesta

La matriz traspuesta es la matriz que se obtiene cambiando entre sí las filas y las columnas de una matriz.

Matriz simétrica

Una matriz es simétrica si es igual a su traspuesta.

Matriz singular

Una matriz es singular si su determinante es igual a 0. Las matrices singulares no son invertibles.

Media muestral

La media muestral es igual a la media aritmética de los valores de una muestra. Es una estimación del valor medio.

Menor complementario

El menor complementario (i, j) de una matriz A es el determinante de la matriz obtenida a partir de A eliminando la fila i y la columna j .

Menor de una matriz

El menor de una matriz es el determinante de una submatriz obtenida eliminando algunas filas y/o columnas.

Método de máxima verosimilitud

El método de máxima verosimilitud es un método de estimación de parámetros que consiste en encontrar los parámetros que maximizan la función de verosimilitud.

Momento de orden n

El momento de orden n de una variable aleatoria se define como $E[X^n] = \int x^n f_x(x) dx$. En particular, el momento de orden 1 es el valor medio.

Ortogonales

Dos vectores son ortogonales si el producto escalar entre ellos es igual a 0.

Preimagen

Véase *imagen inversa*.

Primitiva de una función

F es una primitiva de la función f si $F' = f$. La primitiva es, pues, la inversa de la derivada. Es importante observar que las primitivas de una función no son únicas.

Probabilidad

La probabilidad es un valor numérico positivo que indica cómo de probable es que un evento se realice.

Probabilidad condicionada

La probabilidad condicionada de un evento E dado otro evento F se define como $P(E|F) = P(E \cap F)/P(F)$ y representa la probabilidad de que se realice el evento E si sabemos que el evento F ya se ha realizado.

Producto escalar

El producto escalar es una operación binaria que asocia a cada pareja de vectores la suma de los productos de sus componentes.

Punto crítico

Un punto crítico de una función es un punto de su dominio donde la derivada es 0.

Punto estacionario

Véase *punto crítico*.

Punto de máximo absoluto

Un punto de máximo absoluto de una función es un punto del dominio de la función tal que el valor de la función es el mayor.

Punto de máximo global

Véase *punto de máximo absoluto*.

Punto de mínimo absoluto

Un punto de mínimo absoluto de una función es un punto del dominio de la función tal que el valor de la función es el menor.

Punto de mínimo global

Véase *punto de mínimo absoluto*.

Punto de máximo relativo

Un punto de máximo relativo es un punto del dominio tal que es un punto de máximo en un subconjunto, es decir, un entorno del punto.

Punto de máximo local

Véase *punto de máximo relativo*.

Punto de mínimo relativo

Un punto de mínimo relativo es un punto del dominio tal que es un punto de mínimo en un subconjunto, es decir, un entorno del punto.

Punto de mínimo local

Véase *punto de mínimo relativo*.

Rango de una matriz

El rango de una matriz es la dimensión del más grande menor distinto de 0.

Sistema compatible

Un sistema de ecuaciones es compatible si admite por lo menos una solución.

Sistema incompatible

Un sistema es incompatible si no admite ninguna solución.

Sistema compatible determinado

Un sistema es compatible determinado si tiene una sola solución.

Sistema compatible indeterminado

Un sistema es compatible indeterminado si tiene infinitas soluciones.

Suceso

Véase *evento*.

Suceso elemental

Un suceso elemental es un elemento de un espacio muestral, es decir, un suceso de cardinalidad 1.

Traza

La traza de una matriz cuadrada es la suma de los elementos en la diagonal.

Triangular inferior

Una matriz es triangular inferior si los elementos que hay por encima de la diagonal son iguales a 0.

Triangular superior

Una matriz es triangular superior si los elementos que hay por debajo de la diagonal son iguales a 0.

Valor esperado

Véase *valor medio*.

Valor medio

El valor medio de una variable aleatoria es $E[X] = \int f_x(x)dx$.

Valor propio

Véase *autovalor*.

Valores singulares

Los valores singulares de una matriz (en general, no cuadrada) A son los autovalores de la matriz A^tA .

Variable aleatoria

Una variable aleatoria es una función a partir de un espacio de probabilidad en los números reales.

Variable aleatoria continua

Una variable aleatoria es continua si el conjunto imagen es un conjunto continuo.

Variable aleatoria discreta

Una variable aleatoria es discreta si el conjunto imagen es finito o numerable, es decir, si no es continuo.

Variable aleatoria normal estándar

Una variable aleatoria normal estándar es una variable aleatoria continua con distribución gaussiana (o normal) con media 0 y varianza 1.

Variables independientes e idénticamente distribuidas (i.i.d.)

Las variables aleatorias de un conjunto son independientes e idénticamente distribuidas si son todas independientes entre sí y además tienen todas la misma distribución.

Varianza

La varianza de una variable aleatoria se define como $\text{Var}(X)=E[X^2]-E[X]^2$ y es una medida de la dispersión de la variable alrededor de su valor medio.

Vector

Un vector es un objeto matemático con una dimensión que se representa como una lista ordenada de valores $v=(v_1, v_2, \dots, v_n)$.

Vector columna

Un vector columna es un vector interpretado como una matriz de dimensión $n\times 1$.

Vector fila

Un vector fila es un vector interpretado como una matriz de dimensión $1\times n$.

Vector propio

Véase *autovector*.

Verosimilitud

La función de verosimilitud es el producto de la densidad de la variable aleatoria en los puntos de la muestra considerando los parámetros de la densidad como variables.

Versión estandarizada

La versión estandarizada de una variable aleatoria es una trasformación de la variable obtenida con una traslación y una escala de tal forma que la nueva variable aleatoria tenga valor medio 0 y varianza 1.



Enlaces de interés

The mathematician who cracked Wall Street | Jim Simons

Charla TED con Jim Simons, profesor de matemáticas y fundador de Renaissance Technologies. Fue uno de los primeros en aplicar técnicas matemáticas y estadísticas en la predicción y la optimización en el mercado bursátil, un ejemplo de la aplicación de las matemáticas. En Renaissance Technologies empezaron a utilizar el aprendizaje automático y los datos masivos (big data) para entrenar modelos matemáticos.

<https://www.youtube.com/watch?v=U5kldtMJGc8>

MATHSCINET, Mathematical Reviews

MathSciNet es un repositorio de artículos de matemáticas. Es la mejor elección para encontrar artículos y autores en una fuente oficial. Lo gestiona la American Mathematical Society.

<https://mathscinet.ams.org/mathscinet/>

divulgaMAT

divulgaMAT es el centro virtual de divulgación de las matemáticas de la Real Sociedad Matemática Española.

<http://www.divulgamat.net/>

GeoGebra

Es una herramienta en línea para dibujar gráficas de funciones, realizar construcciones geométricas y resolver ecuaciones. Está disponible en español.

<https://www.geogebra.org/>

Bibliografía

Cohn, P. (1994). *Elements of Linear Algebra* (p. 69). Londres: Taylor & Francis Ltd/CRC Press.

Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*, 2 (11), 559–572. doi:10.1080/14786440109462720.



Autor