Grandes de Bases de Datos

Modelos de clasificación y predicción

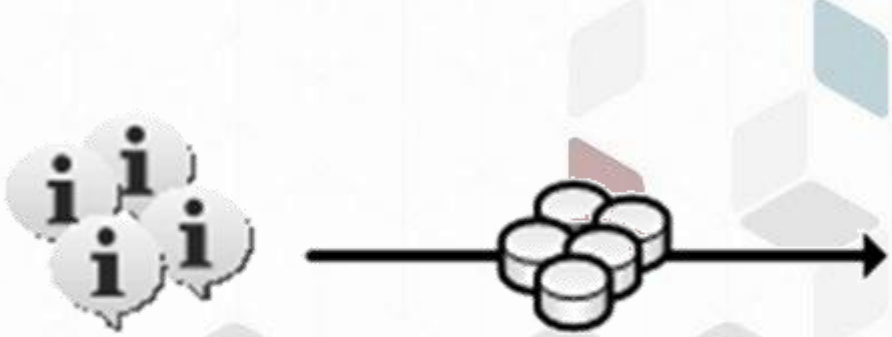
Naïve Bayes

Modelos predictivos
Modelos descriptivos

Aprendizaje supervisado

La clasificación y la predicción pertenecen a este tipo de aprendizaje, dado que se introduce al algoritmo de aprendizaje ejemplos (salidas) reales para que con base en ellos pueda generar un modelo capaz de generalizar los patrones encontrados.

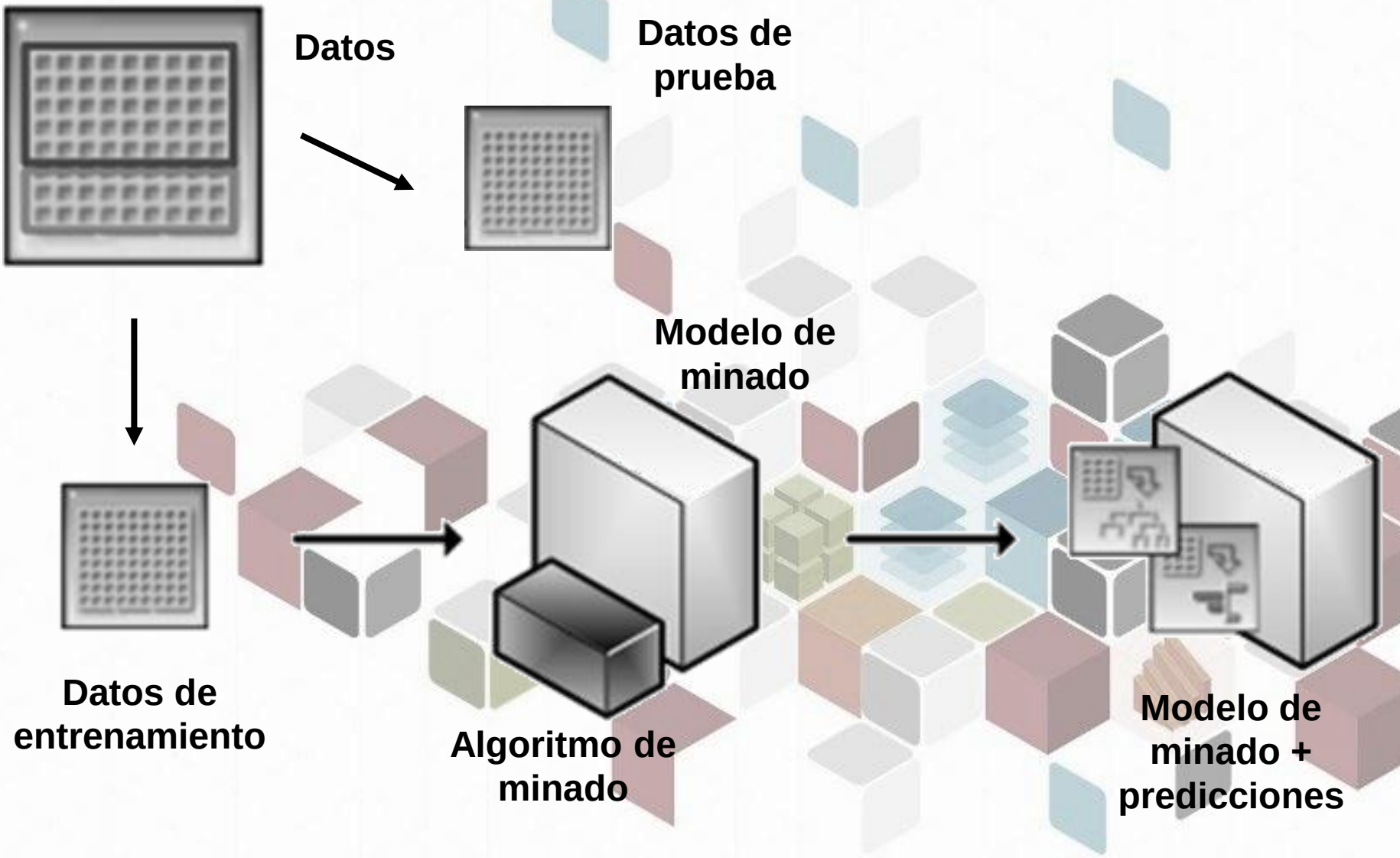
Modelos de Minado



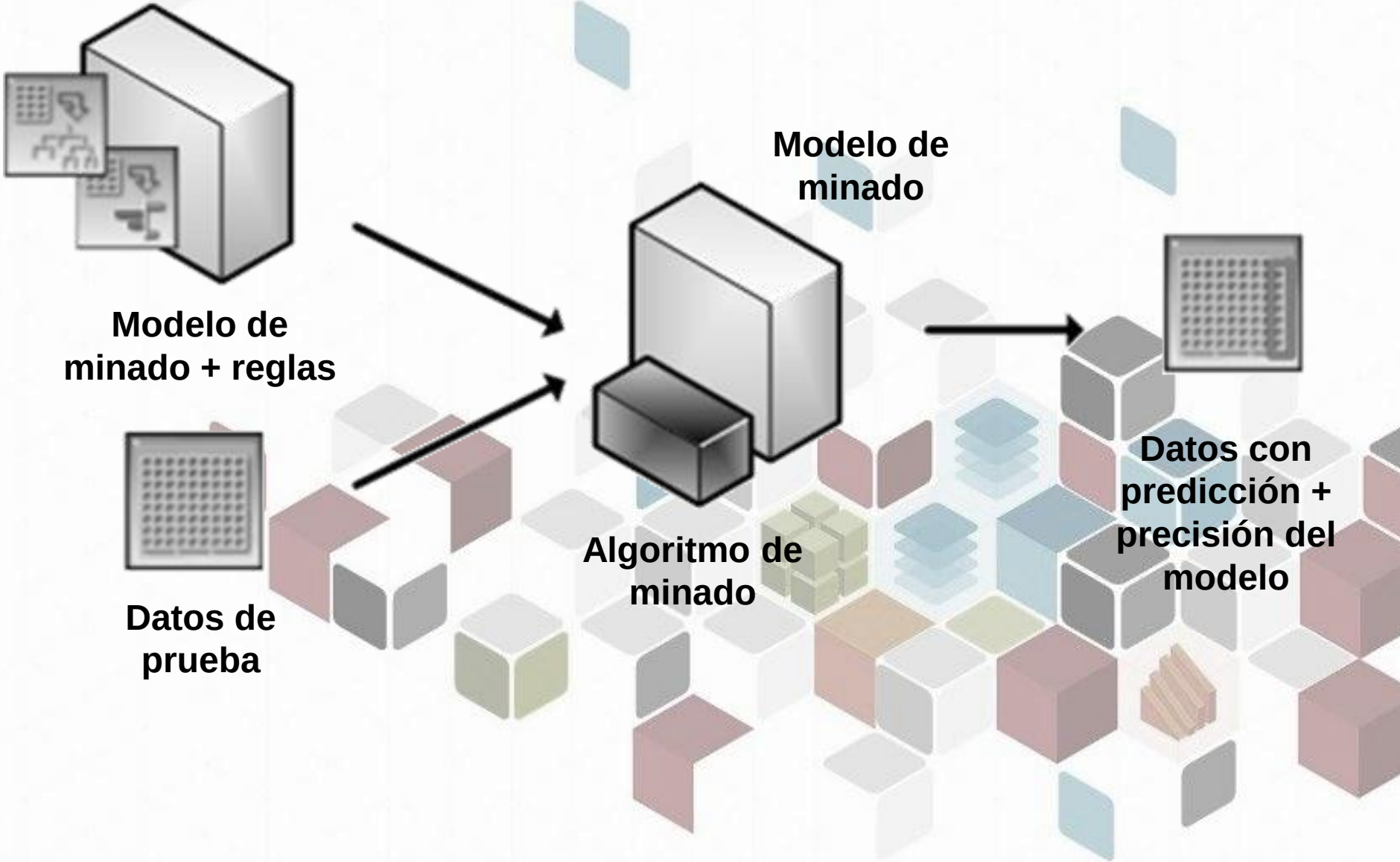
Conocimiento

Modelo de minado

Modelos de Minado



Modelos de Minado



Predicción

Dentro de la minería de datos esta tarea se divide en dos ramas:

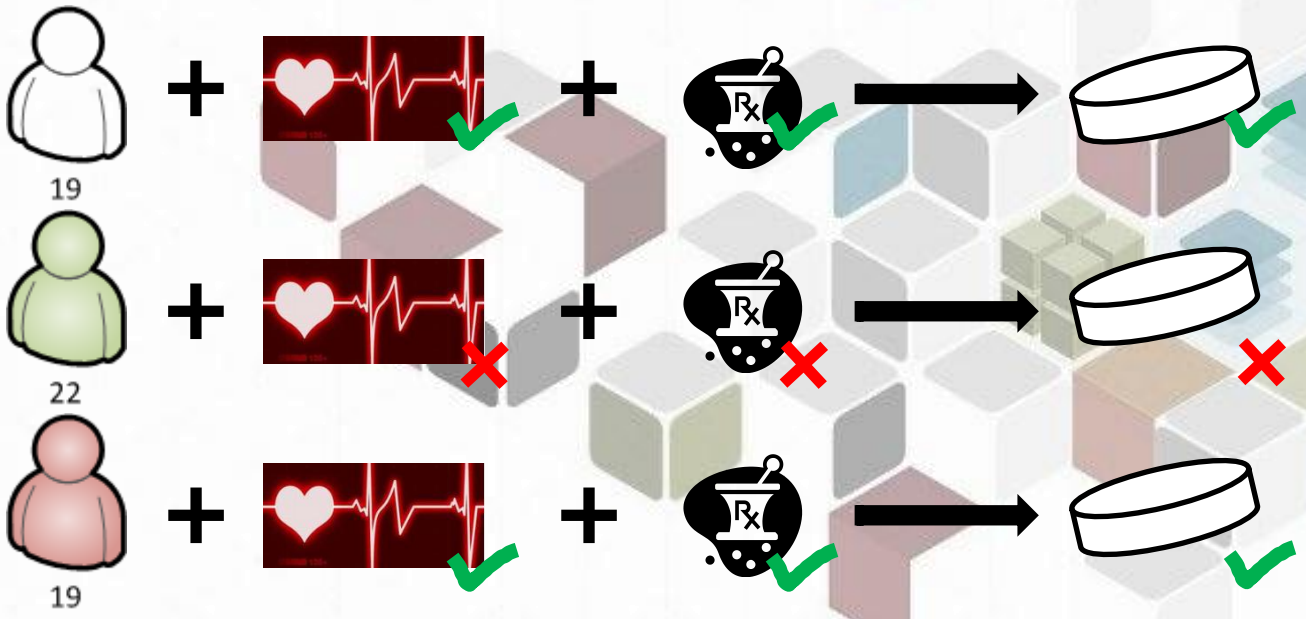
Clasificación

Es quizá la tarea mas utilizada.

- Los datos son instancias caracterizadas por diferentes atributos y pertenecen a una clase (o distintas).
- La tarea consiste en generar un modelo que dado una instancia provea la clase a la que pertenece.

Clasificación

- **Clasificación**
  - Encontrar un modelo para predecir la clasificación de datos no observados (conociendo las etiquetas).



Pacientes menores de 19 años, sin presión alta y análisis Rx, pueden tomar la medicina

Clasificación

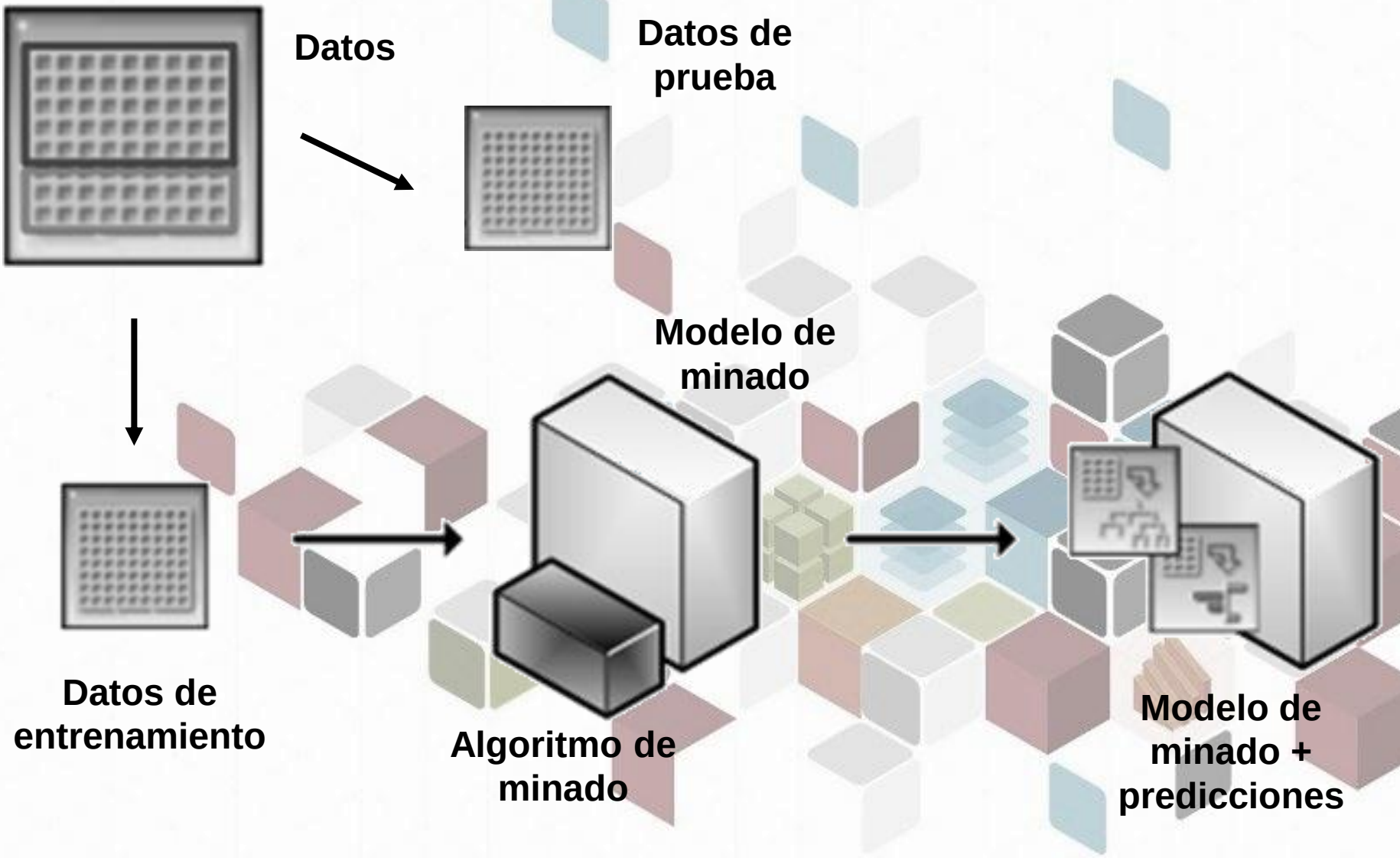
- **Clasificación**
  - Encontrar un modelo para predecir la clasificación de datos no observados (conociendo las etiquetas).
  - Algoritmos de clasificación:
    - Clasificador ingenuo de Bayes.
    - Árboles de clasificación
    - Redes neuronales
    - K-vecinos más próximos

Clasificación

La clasificación es un proceso de dos pasos:

Primer Paso:
Un modelo es construido describiendo un conjunto predeterminado de cases de datos. El modelo es construido analizando tuplas de bases de datos descritas por atributos. Las tuplas analizadas para construir el modelo colectivo forman el conjunto de entrenamiento.

Modelos de Minado

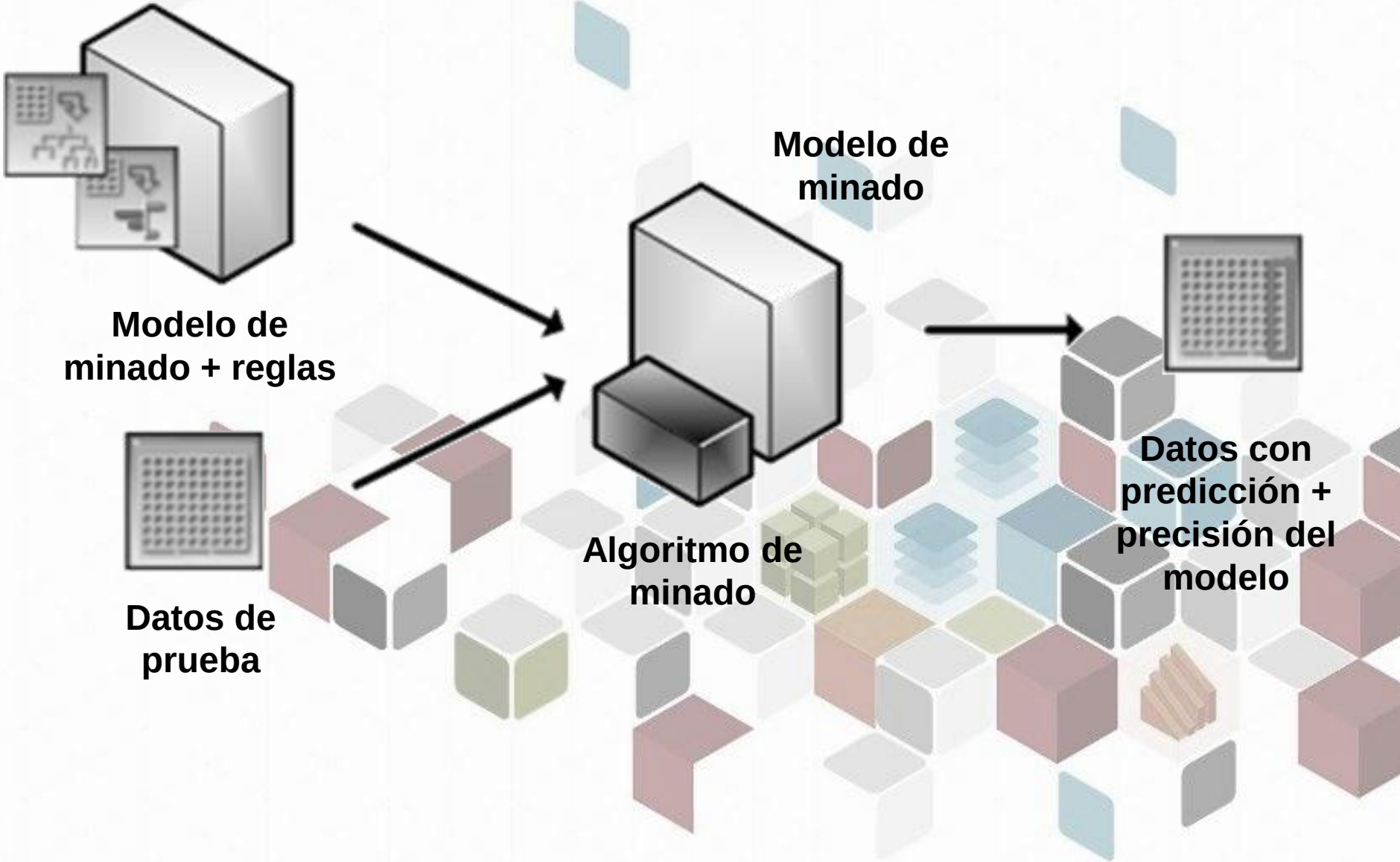


Clasificación

Segundo Paso:

El modelo es usado para clasificación. Se estima la precisión del modelo. Esto se realiza utilizando un conjunto de prueba. Se le presentan al modelo los ejemplos de prueba y se compara su respuesta con la esperada, se calcula el error y si el error es aceptable, entonces el modelo puede ser usado para clasificar datos nuevos y desconocidos.

Modelos de Minado



Conceptos

- Cada atributo (A_1,A_2,…,A_n) representa una característica.
- Una tupla se representa como: X=(x_1,x_2,…,x_n)
- Cada tupla pertenece a alguna clase predefinida la cual denotaremos como: “atributo clasificador”.
- El atributo clasificador es discreto y no ordenado.

Predicción

- La predicción de datos (numéricos) es un proceso de dos pasos similar a la clasificación
- La variable a predecir se denotará como “atributo de predicción” o simplemente “pronóstico”.
- Este tipo de técnicas sirve para resolver preguntas como: “¿Cuánto serán las ventas para el siguiente mes?”.

Preparando datos

- *Limpiado de datos*.
  - Eliminar o reducir el ruido y el tratamiento de valores nulos.
- *Análisis de relevancia*.
  - Identificar atributos importantes para nuestro estudio (ejem. análisis de correlación).
- *Transformación y reducción de datos*.
  - Datos modificados por normalización, por jerarquías, componentes principales, entre otros.

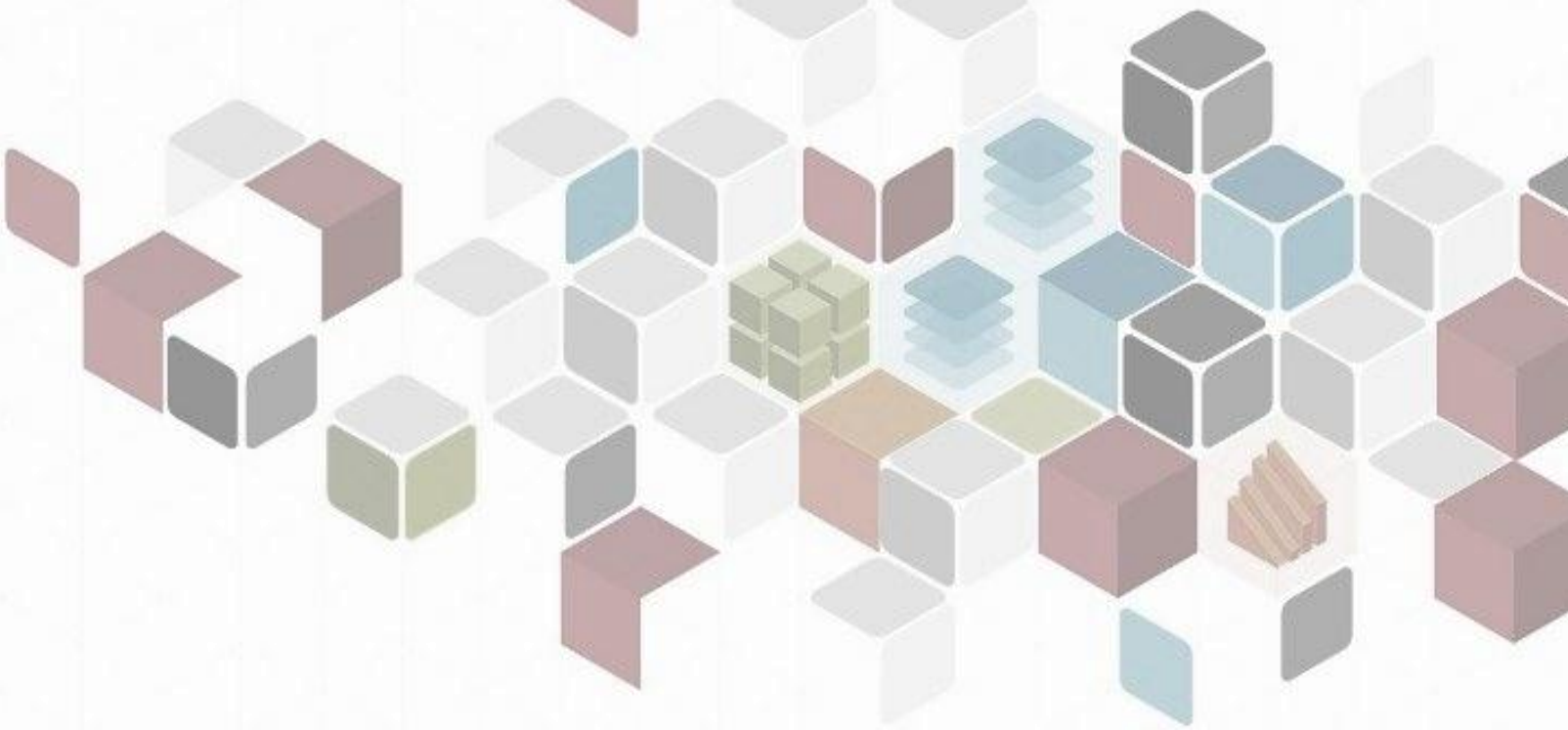
Técnicas y sus usos

Nombre	PREDICTIVO		DESCRIPTIVO		
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlaciones/ Factorizaciones
Redes Neuronales	✓	✓	✓		
Árboles de decisión ID3, C4.5, C5.0	✓				
Árboles de decisión CART	✓	✓			
Otros árboles de decisión	✓	✓	✓	✓	
Redes de Kohonen			✓		
Regresión lineal y logarítmica		✓			✓
Regresión logística	✓			✓	
Kmeans			✓		
Apriori				✓	
Naive Bayes	✓				
Vecinos mas próximos	✓	✓	✓		
Análisis factorial y de comp. ppales.					✓
Twostep, Cobweb			✓		
Algoritmos genéticos y evolutivos	✓	✓	✓	✓	✓
Maquinas de soporte vectorial	✓	✓	✓		
CN2 rules (cobertura)	✓			✓	
Análisis discriminante multivariante	✓				

Tipos de validación

- Validación interna
  - Se aprende, clasifica y valida con los datos de un mismo conjunto
- Validación externa
  - Se aprende un modelo con un conjunto de datos, y se valida con unos datos que no han sido empleados en el aprendizaje

Clasificador Bayesiano



Clasificador Bayesiano

- Son clasificadores estadísticos, que construyen modelos que predicen la probabilidad de posibles resultados (pertenecer a una clase).

- Están basados en el teorema de Bayes.

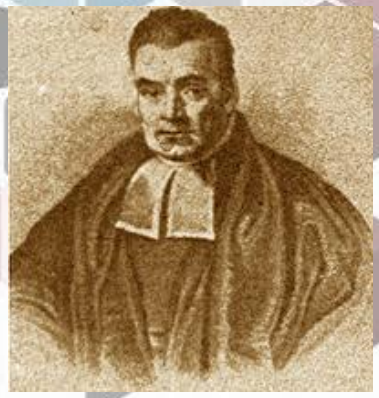
- Son comparables en desempeño a los árboles de decisión y redes neuronales

Teorema de Bayes

Definido por Thomas Bayes en el siglo 18.

Probabilidad de un evento H dada la evidencia X:

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$



Thomas Bayes
1702 - 1761

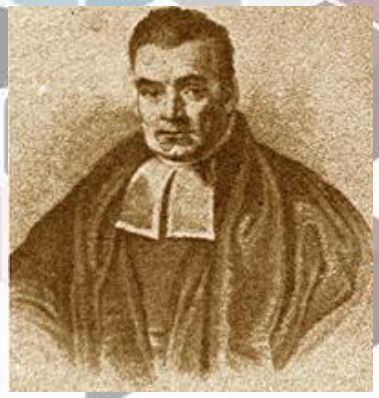
Teorema de Bayes

Definido por Thomas Bayes en el siglo 18.

Probabilidad *A priori* de un evento H:

- Probabilidad del evento, antes de que la evidencia sea conocida

$$P(X)$$



Thomas Bayes
1702 - 1761

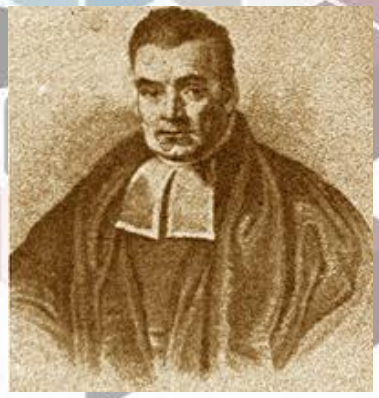
Teorema de Bayes

Definido por Thomas Bayes en el siglo 18.

Probabilidad *A posteriori* de un evento H:

- Probabilidad del evento, después de que la evidencia es conocida

$$P(H \mid X)$$



Thomas Bayes
1702 - 1761

Teorema de Bayes

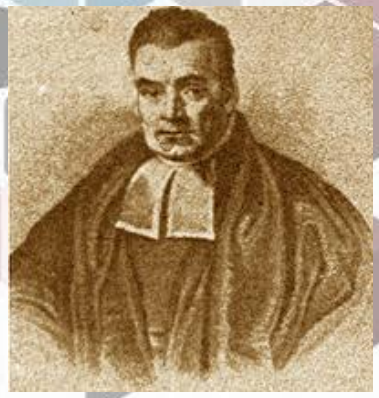
Definido por Thomas Bayes en el siglo 18.

Contexto de clasificación:

$$P(C \mid A_i,\ldots A_n) = \frac{P(A_i,\ldots A_n \mid C)P(C)}{P(A_i,\ldots A_n)}$$

C es una clase

$A_i,\ldots A_n$ conjunto de atributos



Thomas Bayes
1702 - 1761

Teorema de Bayes

Sea ***X*** una tupla de datos.

Sea ***H*** una hipótesis que ***X*** pertenece a una clase ***C***.

P(X|H) es la probabilidad de que dada clase ***C***, la tupla ***X*** pertenezca a ella.

Se desea encontrar ***P(H|X)***; la probabilidad de que dada la tupla ***X*** pertenezca a la clase ***C***

Teorema de Bayes - Ejemplo

Tenemos los atributos edad e ingreso y X es un cliente de 35 años e ingreso de 40,000. Supongamos que H es la hipótesis de que nuestro cliente comprará una computadora.

- ***P(H|X).-*** refleja la probabilidad que el cliente X comprará una computadora dado que conocemos la edad y el ingreso del cliente.

- ***P(H).-*** es la probabilidad a priori de H, es la probabilidad de que cualquier cliente compre una computadora.

- ***P(X).-*** es la probabilidad de que una persona de nuestro conjunto de datos tenga 35 años y gane 40,000.

- ***P(X|H)***.- es la probabilidad de que un cliente de 35 años y que gane 40,000, dado que sabemos que compró una computadora.

Calculando las probabilidades

- *P(H)*, *P(X)* y *P(X|H)* pueden ser estimadas a partir del conjunto de datos.
- *P(H|X)* es calculado a través del teorema de Bayes, definido anteriormente:

$$P(H \mid X) = \frac{P(X \mid H)P(H)}{P(X)}$$

Clasificador Naïve Bayes

Es una simplificación del teorema de Bayes.

El clasificador trabaja como sigue:

1.-Sea *D* un conjunto de entrenamiento con su clase clasificadora asociada.

Clasificador Naïve Bayes

2.-Supongamos que hay *m* clases $C_1, C_2, \ldots C_m$. Dada una *tupla X*, el clasificador predecirá que *X* pertenece a la clase que tiene la probabilidad posteriori mayor, condicionada a *X*. Esto es:

$$P(C_i \mid X) > P(C_j \mid X) \qquad 1 \leq j \leq m, j \neq i.$$

Ojo… esto quiere decir que se debe conocer la probabilidad para todas las clases

Clasificador Naïve Bayes

3.- *P(X)* es constante para todas las clases, entonces sólo $P(X|C_i)P(C_i)$ es necesario ser maximizado. Para estimar las probabilidades de las clases, se utiliza:

$$P(C_i) = \frac{|C_{i,D}|}{|D|}$$

Donde $|C_{i,D}|$ es el número de *tuplas* de entrenamiento de la clase C_i en *D*.

Clasificador Naïve Bayes

4.- Para reducir el costo computacional, Naïve Bayes asume que la evidencia, se divide en partes, es decir, los atributos son independientes. De modo que los cálculos se reducen a:

$$P(X \mid C_i) = \prod_{k=1}^{n} P(x_k \mid C_i) = P(x_1 \mid C_i) \cdot P(x_2 \mid C_i) \cdots \cdots P(x_n \mid C_i)$$

Clasificador Naïve Bayes

- a) Si A_k es categórico, entonces $P(x_k|C_i)$ es el número de tuplas de la clase C_i en D que tienen el valor x_k en A_k. Dividido por $|C_{i,D}|$.

- b) Si A_k es continuo, entonces se asume que los valores siguen una distribución normal con media μ y desviación estándar σ. Definida por:

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

Ejemplo. Sea el atributo edad el cual es continuo. Primero se calcula la media y la desviación estándar, supongamos que fue μ=38 y σ=38±12 y X es 35, entonces aplicando la formula tenemos g(35,38,50)=.003177

Clasificador Naïve Bayes

5.- Por último para predecir la clase clasificadora de *X*, $P(X|C_i)P(C_i)$ es calculado para cada clase C_i. El clasificador predice que la clase clasificadora para la *tupla X* es la clase C_i. Si.

$$P(X \mid C_i)\, P(C_j) > P(X \mid C_j)\, P(C_j)$$

$$1 \leq j \leq m,\ j \neq i.$$

Clasificador Naïve Bayes

Naïve Bayes predice resultados binarios o multiclase.

En los problemas binarios, cada registro cumplirá o no el comportamiento modelado {si, no}.

Puede hacer predicciones para problemas multiclase, en los cuales hay varios resultados posibles {bueno, malo, regular}.

Naïve Bayes - Ejemplo

- tiendaWeb dispone de los siguientes datos sobre sus clientes, clasificados en buenos y malos clientes

nIdCliente	Educacion	EdoCivil	Genero	nEdad	Hijos	BuenCliente
28	Educación secundaria	C	F	Joven	3	Si
29	Estudios de postgrado	C	M	Joven	3	No
30	Educación secundaria	C	M	Joven	3	No
31	Licenciatura	C	M	Joven	2	No
32	Educación secundaria	S	M	Mayor	3	No
33	Licenciatura	C	M	Mayor	0	Si
34	Licenciatura	C	M	Joven	2	No
35	Licenciatura	C	F	Joven	2	Si
36	Estudios de postgrado	C	M	Mayor	3	No
37	Licenciatura	S	F	Joven	2	Si

Naïve Bayes - Ejemplo

nIdCliente	Educacion	EdoCivil	Genero	nEdad	Hijos	BuenCliente
28	Educación secundaria	C	F	Joven	3	Si
29	Estudios de postgrado	C	M	Joven	3	No
30	Educación secundaria	C	M	Joven	3	No
31	Licenciatura	C	M	Joven	2	No
32	Educación secundaria	S	M	Mayor	3	No
33	Licenciatura	C	M	Mayor	0	Si
34	Licenciatura	C	M	Joven	2	No
35	Licenciatura	C	F	Joven	2	Si
36	Estudios de postgrado	C	M	Mayor	3	No
37	Licenciatura	S	F	Joven	2	Si

Suponiendo que “llega” el siguiente registro:

38	Licenciatura	C	M	Mayor	0	***?***

Naïve Bayes - Ejemplo

La hipótesis C es que buen_cliente = sí.
La tupla X es una combinación de los valores de los atributos Educación, EdoCivil, Género, Edad, e Hijos, por lo que su probabilidad se obtiene multiplicando las probabilidades de estos valores. Es decir:

$$P(Si \mid X) = \frac{[P(Educ \mid Lic)P(EdoCivil \mid C)P(Gén \mid M)P(Edad \mid Mayor)P(Hijos \mid 0)]P(Si)}{P(X)}$$

Por Naïve Bayes desaparece P(X)
P(Educ|Lic) se calcula dividiendo el número de instancias que tienen el valor "Licenciatura" en el atributo Educación (de los que en buen_cliente es "Si") dividido por el número de instancias cuyo valor del atributo buen_cliente es "Si".
Es decir *P(Educ|Lic)*= 3/4

Naïve Bayes - Ejemplo

La hipótesis C es que buen_cliente = sí.
La tupla X es una combinación de los valores de los atributos Educación, EdoCivil, Género, Edad, e Hijos, por lo que su probabilidad se obtiene multiplicando las probabilidades de estos valores. Es decir:

$$P(Si \mid X) = \frac{[P(Educ \mid Lic)P(EdoCivil \mid C)P(Gén \mid M)P(Edad \mid Mayor)P(Hijos \mid 0)]P(Si)}{P(X)}$$

P(EdoCivil|C) se calcula dividiendo el número de instancias que tienen el valor "C" en el atributo EdoCivil (de los que en buen_cliente es "Si") dividido por el número de instancias cuyo valor del atributo buen_cliente es "Si".
Es decir *P(EdoCivil|C)*= 3/4

Naïve Bayes - Ejemplo

La hipótesis C es que buen_cliente = sí.
La tupla X es una combinación de los valores de los atributos Educación, EdoCivil, Género, Edad, e Hijos, por lo que su probabilidad se obtiene multiplicando las probabilidades de estos valores. Es decir:

$$P(Si \mid X) = \frac{[P(Educ \mid Lic)P(EdoCivil \mid C)P(Gén \mid M)P(Edad \mid Mayor)P(Hijos \mid 0)]P(Si)}{P(X)}$$

P(Educ|Lic) = 3/4
P(EdoCivil|C) = 3/4
P(Gén|M) = 1/4
P(Edad|Mayor) = 1/4
P(Hijos|0) = 1/4

$$P(Si) = 4/10$$

Naïve Bayes - Ejemplo

La hipótesis C es que buen_cliente = sí.
La tupla X es una combinación de los valores de los atributos Educación, EdoCivil, Género, Edad, e Hijos, por lo que su probabilidad se obtiene multiplicando las probabilidades de estos valores. Es decir:

$$P(Si \mid X) = (\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4}) \frac{4}{10}$$

$$= (.75)(.75)(.25)(.25)(.25)(.40)$$

$$= 0.00351562$$

P(Educ|Lic) = 3/4
P(EdoCivil|C) = 3/4
P(Gén|M) = 1/4
P(Edad|Mayor) = 1/4
P(Hijos|0) = 1/4

$$P(Si) = 4/10$$

Naïve Bayes - Ejemplo

Procediendo de igual manera para la clase “**No**” resulta:

$$P(No/X) = ??$$

Por lo que se asignará el valor **??** al atributo buen_cliente del nuevo dato o tupla

¿Probabilidades iguales a cero?

Para solucionar este problema se suele aplicar una corrección Laplaciana, la cual consiste en sumarle 1 a todos los distintos valores de la clase en ese atributo.

Ejemplo.
Supongamos que la clase BuenCliente = “Si” para D que contiene 1000 tuplas. Tenemos 0 tuplas para Educacion = “Licenciatura”, 990 para Educacion = “Secundaria” y 10 para Educacion = “Postgrado”.

$P(Educ|Lic)$ = 0/1000 = 0
$P(Educ|Sec)$ = 990/1000 = 0.99
$P(Educ|Post)$ = 10/1000 = 0.01

¿Probabilidades iguales a cero?

Para solucionar este problema se suele aplicar una corrección Laplaciana, la cual consiste en sumarle 1 a todos los distintos valores de la clase en ese atributo.

Ejemplo.
Aumentando 1, se tiene lo siguiente

P(Educ|Lic) = 1/1000 = 0.001
P(Educ|Sec) = 991/1000 = 0.991
P(Educ|Post) = 11/1000 = 0.011

Las diferencias son marginales y se evita trabajar con probabilidades igual a 0

¿Valores NULL?

Si existen valores NULL, se tratan según donde aparezca la marca, de la siguiente forma:

- Etapa de entrenamiento:
  - La instancia no se incluye en la obtención de la cuenta de frecuencia para la combinación valor-clase
- Etapa de clasificación:
  - El atributo se omite del calculo

¿Qué tan efectivo es el clasificador Bayesiano?

Varios estudios empíricos han demostrado que el clasificador Bayesiano es comparable en relación con los árboles de decisión y las redes neuronales en algunos dominios. Sin embargo esto dependerá si se cumple la hipótesis sobre la independencia de los atributos (en cuyo caso el desempeño será menor).