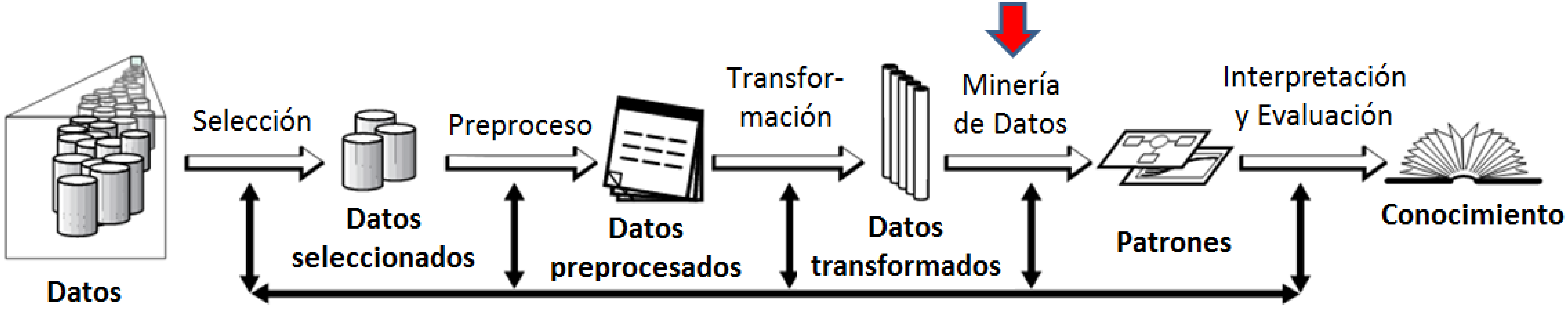
CLASIFICADOR NAÏVE BAYES

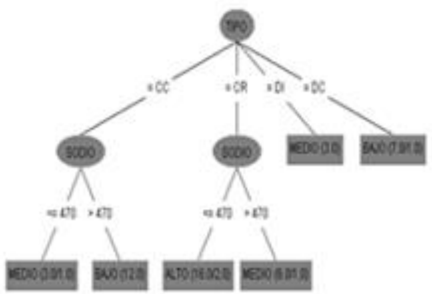
Minería de Datos y el proceso de KDD



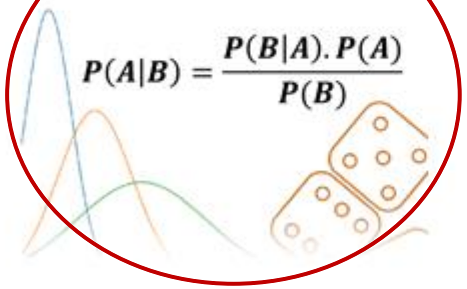
Fayyad (1996)

- **Técnicas de Minería de Datos**

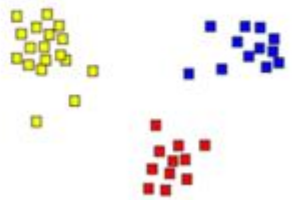
ARBOL



CLASIF. BAYESIANO



AGRUPAMIENTO



REGLAS

IF (TIPO = CC) AND (SODIO > 470)
ENTONCES (COSTO=BAJO)

IF (TIPO = CR) AND (PRODUCTO = CN)
ENTONCES (COSTO=ALTO)

IF (TIPO = DC)
ENTONCES (COSTO=MEDIO)

Clasificador Bayesiano

- Permite tomar una decisión identificando la situación más probable con base en la ocurrencia de eventos.
- Conceptos relacionados
  - Probabilidad condicional
  - Teorema de la multiplicación
  - Teorema de Bayes
  - Hipótesis máxima a posteriori o hipótesis MAP (*maximum a posteriori*)

Probabilidad Condicional

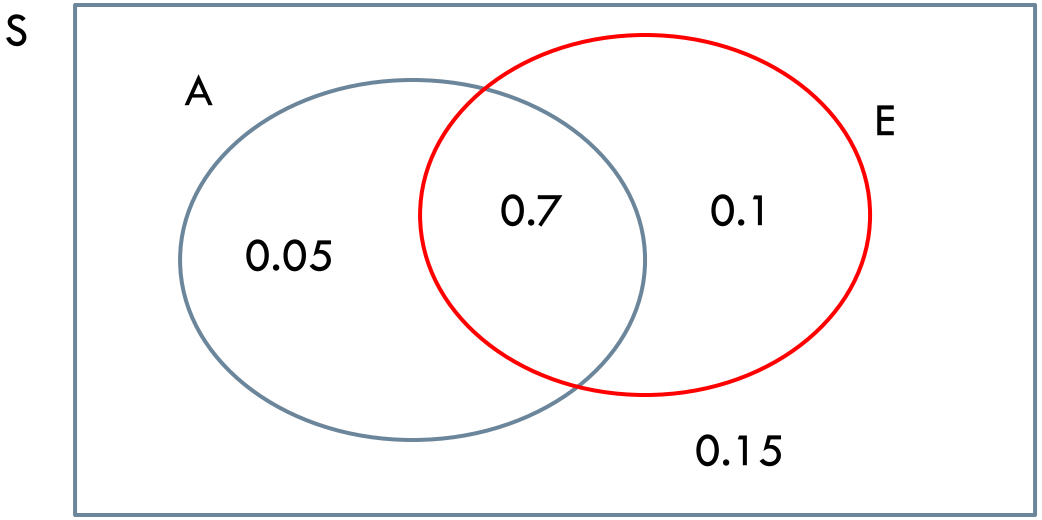
- En ocasiones, resulta necesario calcular la probabilidad de un evento luego de saber que otro ha ocurrido.
- Es decir que la probabilidad del 2do. evento debe calcularse en referencia al espacio muestral determinado por el 1er. evento.

Veamos un ejemplo

Ejemplo

- ε : "Tomar al azar un alumno del curso"
- Eventos : A = "el alumno aprobó el examen"

  E = "el alumno estudió para el examen"



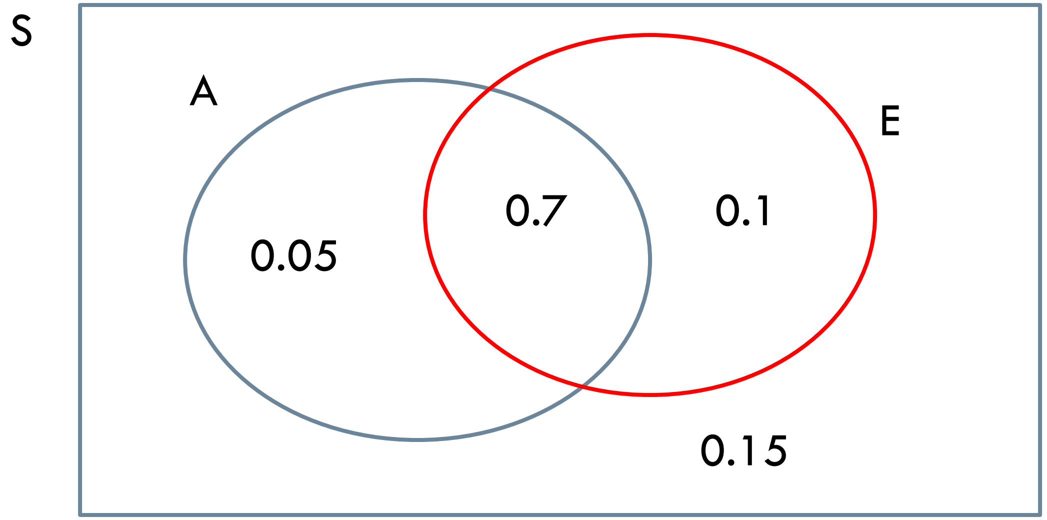
$$P(A) = 0.75$$

$$P(E) = 0.8$$

$$P(A \cap E) = 0.7$$

Probabilidad condicional - Ejemplo

- ¿Cuál sería la probabilidad de que un alumno haya aprobado el examen sabiendo que ha estudiado?



Sabemos que el evento E ocurre, es decir, el alumno ha estudiado.

Intuitivamente

$$P(A\,|E) = \frac{0.7}{0.8} = 0.875$$

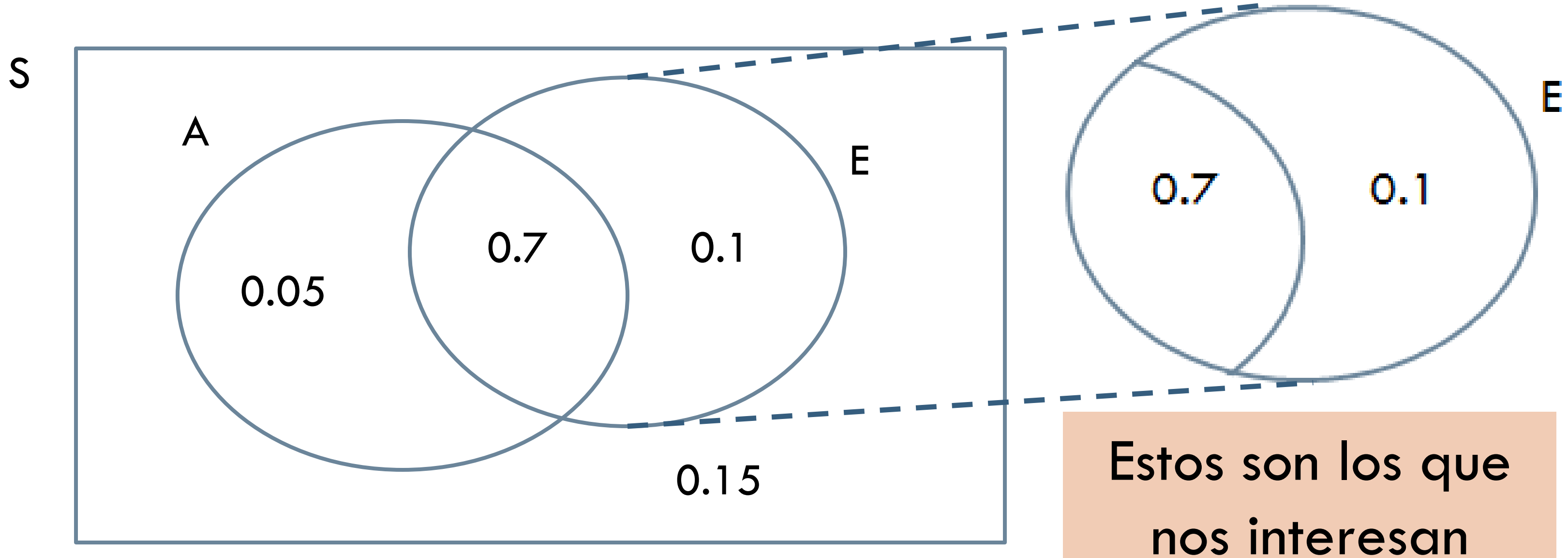
Probabilidad condicional

- Dados dos eventos A y B, con $P(B) > 0$, la probabilidad condicional de A dado que ocurrió B se define como

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probabilidad condicional. Ejemplo

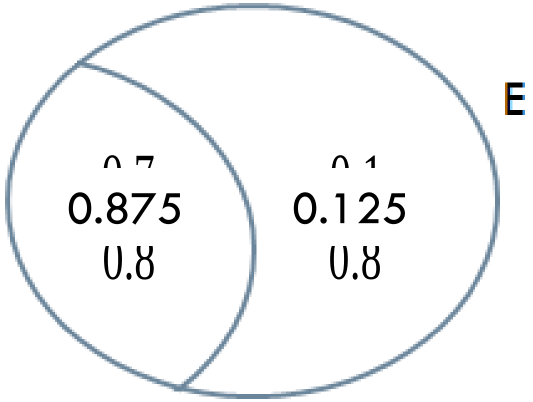
- Al calcular una probabilidad sobre los que estudiaron, se produce un cambio del espacio muestral



Estos son los que nos interesan

Probabilidad condicional

- Estas probabilidades respetan el espacio muestral original.



- Si ahora E es el nuevo espacio muestral, debemos hacer que sumen 1 manteniendo la proporción.
- Es decir que debemos dividirlas por P(E).

Probabilidad condicional

- Estas probabilidades corresponden al nuevo espacio muestral.

Probabilidad de desaprobar habiendo estudiado

Probabilidad de aprobar habiendo estudiado

Ejemplo

- El archivo ***Lentes_Bayes.xls*** contiene 24 muestras correspondientes a diagnósticos de uso de lentes de contactos extraídas de http://archive.ics.uci.edu/ml/datasets/Lenses

- Atributos:
  - Edad del paciente: joven, pre-presbicia, presbicia
  - Prescripción de lentes: Miopía, hipermetropia
  - Astigmatismo: Si, No
  - Producción de lágrimas: Reducida, Normal
  - Diagnóstico: no_usar_lentes, lentes_blandos, lentes_duros

Ejemplo

Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
Joven	Miopía	NO	Reducida	No_usar_Lentes
Joven	Miopía	NO	Normal	Lentes_Blandos
…	…	…	…	…
Joven	Hipermetropía	SI	Reducida	No_usar_Lentes
Joven	Hipermetropía	SI	Normal	Lentes_Duros
pre_presb	Miopía	NO	Reducida	No_usar_Lentes
pre_presb	Miopía	NO	Normal	Lentes_Blandos
…	…	…	…	…
Presbicia	Hipermetropía	SI	Reducida	No_usar_Lentes
Presbicia	Hipermetropía	SI	Normal	No_usar_Lentes

Ejercicio

- ¿Cuál es la probabilidad de que un paciente tenga EDAD = “Joven” si se sabe que su DIAGNOSTICO fue usar “lentes_blandos”?

Id	Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
1	Joven	Hipermetropía	NO	Normal	Lentes_Blandos
2	Joven	Miopía	NO	Normal	Lentes_Blandos
9	pre_presb	Hipermetropía	NO	Normal	Lentes_Blandos
11	pre_presb	Miopía	NO	Normal	Lentes_Blandos
18	Presbicia	Hipermetropía	NO	Normal	Lentes_Blandos

$$P(Edad = \text{"}Joven\text{"} | Diag = \text{"}lentes\_blandos\text{"}) = \frac{P(Edad = \text{"}Joven\text{"} \cap Diag = \text{"}lentes\_blandos\text{"})}{P(Diag = \text{"}lentes\_blandos\text{"})}$$

$$P(Edad = \text{"}Joven\text{"} | Diag = \text{"}lentes\_blandos\text{"}) = \frac{2/24}{5/24} = \frac{2}{5} = 0.4$$

Ejemplo

- ¿Cuál es la probabilidad de que un paciente tenga astigmatismo si se sabe que su producción de lágrimas es reducida?

$$P(astigmatismo = si | lagrimas = reducida) = \frac{P(astigmatismo = si \cap lagrimas = reducida)}{P(lagrimas = reducida)}$$

$$P(astigmatismo = si | lagrimas = reducida) = \frac{6/24}{12/24}$$

$$= \frac{6}{12} = 0.5$$

Evento	# ejemplos
(lagrimas=reducida)	12
(lagrimas = reducida) y (astigatismo=si)	6
Total	**24**

Teorema de la multiplicación

- Usando la definición de probabilidad condicional se obtiene

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

se obtiene

$$\boxed{P(A \cap B) = P(A|B) \, . \, P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

se obtiene

$$\boxed{P(A \cap B) = P(B|A) \, . \, P(A)}$$

Teorema de Bayes

- Sea $\{C_1, C_2, \ldots, C_n\}$ una partición de S
  - $C_1 \cup C_2 \cup \ \ldots \cup C_n = S$
  - $C_i \cap C_j = \emptyset$, si $i \neq j$
  - $P(C_i) > 0 \quad \forall\, i = 1, 2, \ldots, n$
- Para cualquier evento A de S

$$P(C_k \mid A) = \frac{P(A \mid C_k) P(C_k)}{P(A)}$$

- Se conoce
  - La probabilidad de cada elemento de la partición.
  - La probabilidad de A cuando ocurre c/u de los eventos de la partición.

Teorema de Bayes

- Sea $\{C_1, C_2, \dots, C_n\}$ una partición de S
- Para cualquier evento A de S

$$P(C_k \mid A) = \frac{P(A|C_k)P(C_k)}{P(A)}$$

Teorema de Bayes

- $\boldsymbol{S} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3\} = \{(\boldsymbol{diag} = \text{lentes\_blandos}), (\boldsymbol{diag} = \text{lentes\_duros}), (\boldsymbol{diag} = \text{no\_usar\_lentes})\}$

- Para cualquier evento A de S

$$P(C_k \mid A) = \frac{P(A|C_k)P(C_k)}{P(A)}$$

$C_k = (diag = \text{lentes\_blandos})$

$A = (Edad = Joven)$

Teorema de Bayes

- $\boldsymbol{S} = \{\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3\} = \{(\boldsymbol{diag} = \textbf{lentes\_blandos}), (\boldsymbol{diag} = \textbf{lentes\_duros}), (\boldsymbol{diag} = \textbf{no\_usar\_lentes})\}$

- Para cualquier evento A de S

$$P(C_k \mid A) = \frac{P(A|C_k)P(C_k)}{P(A)}$$

$C_k = (diag = \text{lentes\_blandos})$

$A = (Edad = Joven)$

- Ejemplo

$$P(diag = \text{lentes\_blandos}|\text{Edad=joven}) = \frac{P(Edad = Joven|diag = \text{lentes\_blandos})P(diag = \text{lentes\_blandos})}{P(Edad = Joven)}$$

Teorema de Bayes

- $\boldsymbol{S} = \{\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}\} = \{(\boldsymbol{diag} = \textbf{lentes\_blandos}), (\boldsymbol{diag} = \textbf{lentes\_duros}), (\boldsymbol{diag} = \textbf{no\_usar\_lentes})\}$

- Para cualquier evento A de S

$$P(C_k \mid A) = \frac{P(A|C_k)P(C_k)}{P(A)}$$

$$C_k = (diag = \text{lentes\_blandos})$$

$$A = (Edad = Joven)$$

- Ejemplo

$$P(diag = \text{lentes\_blandos|Edad=joven}) = \frac{P(Edad = Joven|diag = \text{lentes\_blandos})P(diag = \text{lentes\_blandos})}{P(Edad = Joven)}$$

$$P(\text{lentes\_blandos} \mid \text{Joven}) = \frac{P(Joven \mid \text{lentes\_blandos}) \ * \ P(\text{lentes\_blandos})}{P(Joven)}$$

Teorema de Bayes

- Tratándose de un problema de clasificación, con una variable clase (C) y un conjunto de variables predictoras o atributos $\{A_1, \dots, A_n\}$, el teorema de Bayes tendría la siguiente forma:

$$P(C = c_i | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_i) P(C = c_i)}{P(A_1 = a_1, \dots, A_n = a_n)}$$

Teorema de Bayes

- Tratándose de un problema de clasificación, con una variable clase (C) y un conjunto de variables predictoras o atributos $\{A_1, \ldots, A_n\}$, el teorema de Bayes tendría la siguiente forma:

$$P(C = c_i | A_1 = a_1, \ldots, A_n = a_n) = \frac{\boxed{P(A_1 = a_1, \ldots, A_n = a_n | C = c_i)} P(C = c_i)}{P(A_1 = a_1, \ldots, A_n = a_n)}$$

- Si consideramos que los eventos $(A_i = a_i)$ son independientes entre si

$$P(A_1 = a_1, \ldots, A_n = a_n | C = c_i) = P(A_1 = a_1 | C = c_i) * P(A_2 = a_2 | C = c_i) * \cdots * P(A_n = a_n | C = c_i)$$

Teorema de Bayes

- Tratándose de un problema de clasificación, con una variable clase (C) y un conjunto de variables predictoras o atributos $\{A_1, \ldots, A_n\}$, el teorema de Bayes tendría la siguiente forma:

$$P(C = c_i | A_1 = a_1, \ldots, A_n = a_n) = \frac{P(A_1 = a_1, \ldots, A_n = a_n | C = c_i) P(C = c_i)}{P(A_1 = a_1, \ldots, A_n = a_n)}$$

- Si C tiene k posibles valores $\{c_1, \ldots, c_k\}$, *interesará determinar el más probable.*

Teorema de Bayes

- Si es un problema de clasificación, con una variable clase $C = \{c_1, \dots, c_k\}$ y un conjunto de atributos $\{A_1, \dots, A_n\}$, el teorema de Bayes recomendará el más probable

$$P(C = c_1 | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_1) P(C = c_1)}{P(A_1 = a_1, \dots, A_n = a_n)}$$

$$P(C = c_2 | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_2) P(C = c_2)}{P(A_1 = a_1, \dots, A_n = a_n)}$$

$$\dots$$

$$P(C = c_k | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_k) P(C = c_k)}{P(A_1 = a_1, \dots, A_n = a_n)}$$

Teorema de Bayes

Se puede eliminar porque sería el mismo para todos los valores de clase.

- Si es un problema de clasificación, con una variable clase $C = \{c_1, \dots, c_k\}$ y un conjunto de atributos $\{A_1, \dots, A_n\}$, el teorema de Bayes recomendará el más probable

$$P(C = c_1 | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_1) P(C = c_1)}{\cancel{P(A_1 = a_1, \dots, A_n = a_n)}}$$

$$P(C = c_2 | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_2) P(C = c_2)}{\cancel{P(A_1 = a_1, \dots, A_n = a_n)}}$$

…

$$P(C = c_k | A_1 = a_1, \dots, A_n = a_n) = \frac{P(A_1 = a_1, \dots, A_n = a_n | C = c_k) P(C = c_k)}{\cancel{P(A_1 = a_1, \dots, A_n = a_n)}}$$

Teorema de Bayes

- Si es un problema de clasificación, con una variable clase $C = \{c_1, \dots, c_k\}$ y un conjunto de atributos $\{A_1, \dots, A_n\}$, el teorema de Bayes recomendará el más probable

$$P(C = c_1 | A_1 = a_1, \dots, A_n = a_n) = P(A_1 = a_1, \dots, A_n = a_n | C = c_1) P(C = c_1)$$

$$P(C = c_2 | A_1 = a_1, \dots, A_n = a_n) = P(A_1 = a_1, \dots, A_n = a_n | C = c_2) P(C = c_2)$$

$$\dots$$

$$P(C = c_k | A_1 = a_1, \dots, A_n = a_n) = P(A_1 = a_1, \dots, A_n = a_n | C = c_k) P(C = c_k)$$

Teorema de Bayes

- Si es un problema de clasificación, con una variable clase $C = \{c_1, \dots, c_k\}$ y un conjunto de atributos $\{A_1, \dots, A_n\}$, para un conjunto de eventos dados $\{A_1 = a_1, \dots, A_n = a_n\}$ el teorema de Bayes recomendará el más probable

$$P(C = c_1 | A_1 = a_1, \dots, A_n = a_n) = P(A_1 = a_1, | C = c_1) * \cdots * P(A_n = a_n | C = c_1) P(C = c_1)$$

$$P(C = c_2 | A_1 = a_1, \dots, A_n = a_n) = P(A_1 = a_1 | C = c_2) * \cdots * P(A_n = a_n | C = c_2) P(C = c_2)$$

$$\dots$$

$$P(C = c_k | A_1 = a_1, \dots, A_n = a_n) = P(A_1 = a_1 | C = c_k) * \cdots * P(A_n = a_n | C = c_k) P(C = c_k)$$

Luego se elige el valor de clase $\boldsymbol{c_k}$ *que tenga la mayor probabilidad de ocurrir*

Hipótesis MAP (máxima a posteriori)

- El valor de clase a devolver será

$$C_{MAP} = \underset{c\, \epsilon\, \Omega_C}{\arg max}\ P(c|a_1, \ldots a_n) = \underset{c\, \epsilon\, \Omega_C}{\arg\max} \frac{P(a_1, \ldots a_n|c)P(c)}{\boxed{P(a_1, \ldots a_n)}}$$

Se puede eliminar porque sería el mismo para todos los valores de clase.

$$C_{MAP} = \underset{c\, \epsilon\, \Omega_C}{\arg max}\, P(a_1, \ldots, a_n\,|c)\ P(c)$$

donde Ω_C representa el conjunto de valores que puede tomar la variable C.

Clasificador Naïve Bayes

- Utiliza la hipótesis MAP suponiendo que, para un valor de clase dado, los atributos son independientes.

$$C_{MAP} = \underset{c \,\epsilon\, \Omega_C}{\arg max}\, P(a_1, \dots, a_n\,|c)\; P(c)$$

$$C_{MAP} = \underset{c \,\epsilon\, \Omega_C}{\arg max}\, P(a_1|c)P(a_2|c) \dots P(a_n|c)\; P(c)$$

$$C_{MAP} = \underset{c \,\epsilon\, \Omega_C}{\arg \max}\, P(c) \prod_{i=1} P(a_i|c)$$

Por tanto, los parámetros que hay que estimar son $P(a_i|c)$ para cada atributo y la probabilidad a priori de la variable clase $P(c)$.

Ejemplo

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

- Se desea utilizar un clasificador Naive Bayes para predecir el valor del atributo C para el siguiente ejemplo: (A1=b ; A2=3.5)

- Se debe calcular

  $P(C=+|A1=b, A2=3.5)$

  $P(C=-|A1=b, A2=3.5)$

  y elegir la de mayor valor

Ejemplo

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

- Se desea utilizar un clasificador Naive Bayes para predecir el valor del atributo C para el siguiente ejemplo: (A1=b ; A2=3.5)

- Se debe calcular

$$P(C=+|A1=b, A2=3.5) = P(A1=b, A2=3.5|C=+) * P(C=+)$$

$$P(C=-|A1=b, A2=3.5)$$

Ejemplo

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

- Se desea utilizar un clasificador Naive Bayes para predecir el valor del atributo C para el siguiente ejemplo: (A1=b ; A2=3.5)

- Se debe calcular

Considerando que los eventos (A1=b) y (A2=3.5) son independientes

$$P(C=+|A1=b, A2=3.5) = P(A1=b|C=+)*P(A2=3.5|C=+) * P(C=+)$$

$$P(C=-|A1=b, A2=3.5)$$

Ejemplo

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

- Se desea utilizar un clasificador Naive Bayes para predecir el valor del atributo C para el siguiente ejemplo: (A1=b ; A2=3.5)

- Se debe calcular

$$P(C=+|A1=b, A2=3.5) = P(A1=b|C=+)*P(A2=3.5|C=+) * P(C=+)$$

$$P(C=-|A1=b, A2=3.5) = P(A1=b|C=-)*P(A2=3.5|C=-) * P(C=-)$$

Ejemplo

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

- Se desea utilizar un clasificador Naive Bayes para predecir el valor del atributo C para el siguiente ejemplo: (A1=b ; A2=3.5)

- Se debe calcular

$$P(C=+|A1=b, A2=3.5) = P(A1=b|C=+)*P(A2=3.5|C=+) * P(C=+)$$

$$P(C=-|A1=b, A2=3.5) = P(A1=b|C=-)*P(A2=3.5|C=-) * P(C=-)$$

Hay que estimar $P(C)$, $P(A1|C)$ y $P(A2|C)$

Estimación de la probabilidad condicional

- Si el **atributo es discreto** la estimación de la probabilidad condicional se basa en las frecuencias relativas.
- Sea M un subconjunto de registros de la BBDD original. Si llamamos $n(a, M)$ al número de elementos de M en los que el atributo A toma el valor a:

$$P(a|M) = \frac{n(a, M)}{n(M)}$$

En nuestro problema de clasificación el subconjunto M estará conformado por los registros que cumplan con el valor de clase indicado

Esta técnica se conoce como ***estimación por máxima verosimilitud***

<u>Desventajas</u>

- Necesita una muestra de gran tamaño
- Sobreajusta a los datos

Estimación de la probabilidad condicional

Estimación por máxima verosimilitud

$$P(a|M) = \frac{n(a,M)}{n(M)}$$

- Ejemplo

$$P(Edad = pre\_preb|diag = Lentes\_blandos) = \frac{2}{5}$$

Id	Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
1	Joven	Hipermetropía	NO	Normal	Lentes_Blandos
2	Joven	Miopía	NO	Normal	Lentes_Blandos
9	pre_presb	Hipermetropía	NO	Normal	Lentes_Blandos
11	pre_presb	Miopía	NO	Normal	Lentes_Blandos
18	Presbicia	Hipermetropía	NO	Normal	Lentes_Blandos

M *está formado por estos 5 casos*

Estimación de la probabilidad condicional

Estimación por máxima verosimilitud

- Ejemplo

$$P(a|M) = \frac{n(a,M)}{n(M)}$$

$$P(Astigmatismo = SI|diag = Lentes\_blandos) = \frac{0}{5} = 0$$

Id	Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
1	Joven	Hipermetropía	NO	Normal	Lentes_Blandos
2	Joven	Miopía	NO	Normal	Lentes_Blandos
9	pre_presb	Hipermetropía	NO	Normal	Lentes_Blandos
11	pre_presb	Miopía	NO	Normal	Lentes_Blandos
18	Presbicia	Hipermetropía	NO	Normal	Lentes_Blandos

***M** está formado por estos 5 casos*

Estimación de la probabilidad condicional

- Cuando se trabaja con pocos registros, el **estimador de Laplace** puede resultar más adecuado

$$P(a|M) = \frac{n(a,M) + 1}{n(M) + |\Omega|}$$

Probabilidad uniforme a priori

es decir que calcula el cociente entre el número de casos favorables más uno dividido por el número de casos totales más el número de valores posibles del atributo.

Estimación de la probabilidad condicional

Estimación usando Laplace

$$P(a|M) = \frac{n(a,M)+1}{n(M)+|\Omega_A|}$$

- Ejemplo

$$P(Edad = pre\_preb|diag = Lentes\_blandos) = \frac{2+1}{5+3} = \frac{3}{8} = 0.375$$

Id	Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
1	Joven	Hipermetropía	NO	Normal	Lentes_Blandos
2	Joven	Miopía	NO	Normal	Lentes_Blandos
9	pre_presb	Hipermetropía	NO	Normal	Lentes_Blandos
11	pre_presb	Miopía	NO	Normal	Lentes_Blandos
18	Presbicia	Hipermetropía	NO	Normal	Lentes_Blandos

***M** está formado por estos 5 casos*

Estimación de la probabilidad condicional

Estimación usando Laplace

- Ejemplo

$$P(a|M) = \frac{n(a,M)+1}{n(M)+|\Omega_A|}$$

$$P(Astigmatismo = SI | diag = Lentes\_blandos) = \frac{0+1}{5+2} = \frac{1}{7} = 0.143$$

Id	Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
1	Joven	Hipermetropía	NO	Normal	Lentes_Blandos
2	Joven	Miopía	NO	Normal	Lentes_Blandos
9	pre_presb	Hipermetropía	NO	Normal	Lentes_Blandos
11	pre_presb	Miopía	NO	Normal	Lentes_Blandos
18	Presbicia	Hipermetropía	NO	Normal	Lentes_Blandos

***M** está formado por estos 5 casos*

Estimación de la probabilidad condicional

- Si se trata de un **atributo continuo**, el clasificador Naïve Bayes supone que dicho atributo sigue una distribución normal.
- Por tanto, sólo hay que calcular, a partir de la BBDD, la media μ y la desviación típica σ condicionadas a cada valor de la variable clase.

$$P(a_i|c) \propto N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}.\sigma} exp\left(-\frac{(a_i - \mu)^2}{2\sigma^2}\right)$$

- Esta estimación tiene el inconveniente de que los datos no siempre siguen una distribución normal.

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que estimar $P(C)$, $P(A1|C)$ y $P(A2|C)$.

P(C)	
+	0,5
-	0,5

P(A2|C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

P(A1|C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1|C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que estimar *P*(C),

$$\mu = \frac{5+4+2}{3} = 3.67$$

P(C)	
+	0,5
-	0,5

P(A2 | C)	
C=+	$N(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$N(\mu=2{,}33,\ \sigma=0{,}61)$

P(A1 | C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1 | C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que (

$$\sigma = \sqrt{\frac{(5-3.67)^2+(4-3.67)^2+(2-3.67)^2}{3-1}} = 1.53$$

P(C)	
+	0,5
-	0,5

P(A2 | C)	
C=+	$N(\mu=3{,}67, \sigma=1{,}53)$
C=-	$N(\mu=2{,}33, \sigma=0{,}61)$

P(A1 | C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1 | C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que estimar *P*(C), *P*(*A1*|C) y *P*(*A2*|C).

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

P(C)	
+	0,5
-	0,5

P(A2|C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

P(A1|C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1|C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

$$P(A1 = a|C = +) = \frac{1}{3} = 0.33$$

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que estimar *P*(C), *P*(*A1*|C) y *P*(*A2*|C).

P(C)	
+	0,5
-	0,5

P(A2|C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

P(A1|C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1|C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

$$P(A1 = b|C = -) = \frac{0}{3} = 0$$

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que estimar $P(C)$, $P(A1|C)$ y $P(A2|C)$.

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

P(C)	
+	0,5
-	0,5

P(A2|C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

P(A1|C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1|C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

$$P(A1 = a|C = +) = \frac{1+1}{3+2} = \frac{2}{5} = 0.4$$

Ejemplo 1: Construir un clasificador Naive Bayes para predecir el valor del atributo C.

- Hay que estimar *P*(C), *P*(*A1*|C) y *P*(*A2*|C).

A1	A2	C
a	5	+
a	2.2	-
a	1.8	-
b	4	+
b	2	+
a	3	-

P(C)	
+	0,5
-	0,5

P(A2|C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

P(A1|C)	+	-
a	0,33	1
b	0,67	0

Máxima verosimilitud

P(A1|C)	+	-
a	0,4	0,8
b	0,6	0,2

Laplace

$$P(A1 = b|C = -) = \frac{0+1}{3+2} = \frac{1}{5} = 0.2$$

Ejemplo 1

- Usando las estimaciones anteriores clasificar el siguiente caso

(A1=b; A2=3.5)

$P(+|(b, 3.5)) =$

$P(-|(b, 3.5)) =$

- Una vez obtenidas las probabilidades condicionales anteriores se elige la mayor.

Ejemplo 1

- Usando las estimaciones anteriores clasificar el siguiente caso

(A1=b; A2=3.5)

$$P(+|(b,\ 3.5)) = P(+) * P(b|+)*N(3.67,1.53 : 3.5)=$$

$$P(-|(b,\ 3.5)) = P(-) * P(b|-)*N(2.33,0.61 : 3.5)=$$

P(C)	
+	0,5
-	0,5

P(A1|C)	+	-
a	0,33	1
b	0,67	0

P(A2|C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

Ejemplo 1

- Usando las estimaciones anteriores clasificar el siguiente caso

  (A1=b; A2=3.5)

$P(+|(b, 3.5)) = P(+) * P(b|+)*N(3.67,1.53 : 3.5)=0.5*0.67*0.26=0.078$

$P(-|(b, 3.5)) = P(-) * P(b|-)*N(2.33,0.61 : 3.5)=0.5*0*0.1 = 0$

P(C)	
+	0,5
-	0,5

P(A1|C)	+	-
a	0,33	1
b	0,67	0

P(A2|C)	
C=+	$N(\mu=3{,}67, \sigma=1{,}53)$
C=-	$N(\mu=2{,}33, \sigma=0{,}61)$

Ejemplo 1

- Usando las estimaciones anteriores clasificar el siguiente caso

  (A1=b; A2=3.5)

$P(+|(b,\ 3.5)) = P(+) * P(b|+)*N(3.67, 1.53 : 3.5) = 0.5*0.67*0.26 = 0.078$

$P(-|(b,\ 3.5)) = P(-) * P(b|-)*N(2.33, 0.61 : 3.5) = 0.5*0.2*0.1 = 0.01$

P(C)	
+	0,5
-	0,5

P(A1 | C)	+	-
a	0,4	0,8
b	0,6	0,2

P(A2 | C)	
C=+	$\mathcal{N}(\mu=3{,}67,\ \sigma=1{,}53)$
C=-	$\mathcal{N}(\mu=2{,}33,\ \sigma=0{,}61)$

Ejemplo 1

- Usando las estimaciones anteriores clasificar el siguiente caso

(A1=b; A2=3.5)

$P(+|(b, 3.5)) = P(+) * P(b|+)*N(3.67,1.53 : 3.5)=0.5*0.6*0.26 = 0.078$

$P(-|(b, 3.5)) = P(-) * P(b|-)*N(2.33,0.61 : 3.5)=0.5*0.2*0.1 = 0.01$

- Normalizando se obtiene que

$P(+|(b, 3.5))=0.89$ y $P(-|(b, 3.5))=0,11$

- Por lo tanto el nuevo caso se clasificará como +

Ejemplo

- El archivo ***Lentes.csv*** contiene 24 muestras correspondientes a diagnósticos de uso de lentes de contactos extraídas de http://archive.ics.uci.edu/ml/datasets/Lenses
- Atributos:
  - Edad del paciente: joven, pre-presbicia, presbicia
  - Prescripción de lentes: Miopía, hipermetropia
  - Astigmatismo: Si, No
  - Producción de lágrimas: Reducida, Normal
  - Diagnóstico: no_usar_lentes, lentes_blandos, lentes_duros

Ejemplo

Edad	Prescripcion	Astigmatismo	Lagrimas	Diagnostico
Joven	Miopía	NO	Reducida	No_usar_Lentes
Joven	Miopía	NO	Normal	Lentes_Blandos
…	…	…	…	…
Joven	Hipermetropía	SI	Reducida	No_usar_Lentes
Joven	Hipermetropía	SI	Normal	Lentes_Duros
pre_presb	Miopía	NO	Reducida	No_usar_Lentes
pre_presb	Miopía	NO	Normal	Lentes_Blandos
…	…	…	…	…
Presbicia	Hipermetropía	SI	Reducida	No_usar_Lentes
Presbicia	Hipermetropía	SI	Normal	No_usar_Lentes

Attribute	Parameter	Lentes_Blandos	Lentes_Duros	No_usar_Lentes
Edad	value=Joven	0.400	0.500	0.267
Edad	value=pre_presb	0.400	0.250	0.333
Edad	value=Presbicia	0.200	0.250	0.400
Edad	value=unknown	0	0	0
Prescripcion	value=Hipermetropía	0.600	0.250	0.533
Prescripcion	value=Miopía	0.400	0.750	0.467
Prescripcion	value=unknown	0	0	0
Astigmatismo	value=NO	1	0	0.467
Astigmatismo	value=SI	0	1	0.533
Astigmatismo	value=unknown	0	0	0
Lagrimas	value=Normal	1	1	0.200
Lagrimas	value=Reducida	0	0	0.800
Lagrimas	value=unknown	0	0	0

Probabilidades condicionales SIN usar LAPLACE

Ejemplo

- ¿Cuál es la probabilidad de que un paciente no use lentes si se sabe que su edad es JOVEN, la prescripción es MIOPIA, no tiene astigmatismo y su producción de lágrimas es reducida?

P(no_usar_lentes|(Joven, miopía, NO, reducidas)=

P(no_usar_lentes) * P(edad=joven|no_usar_lentes)

*** P(prescripción=miopía|no_usar_lentes)**

*** P(astigmatismo=no | no_usar_lentes)**

*** P(lagrimas=reducidas|no_usar_lentes)**

= 0.632 * 0.267 * 0.467 * 0.467 * 0.8

Attribute	Parameter	Lentes_Blandos	Lentes_Duros	No_usar_Lentes
Edad	value=Joven	0.395	0.490	0.266
Edad	value=pre_presb	0.395	0.250	0.332
Edad	value=Presbicia	0.202	0.250	0.398
Edad	value=unknown	0.008	0.010	0.003
Prescripcion	value=Hipermetropía	0.593	0.253	0.532
Prescripcion	value=Miopía	0.398	0.737	0.466
Prescripcion	value=unknown	0.008	0.010	0.003
Astigmatismo	value=NO	0.984	0.010	0.466
Astigmatismo	value=SI	0.008	0.980	0.532
Astigmatismo	value=unknown	0.008	0.010	0.003
Lagrimas	value=Normal	0.984	0.980	0.201
Lagrimas	value=Reducida	0.008	0.010	0.796
Lagrimas	value=unknown	0.008	0.010	0.003

Probabilidades condicionales usando LAPLACE

Clasificación de flores de Iris

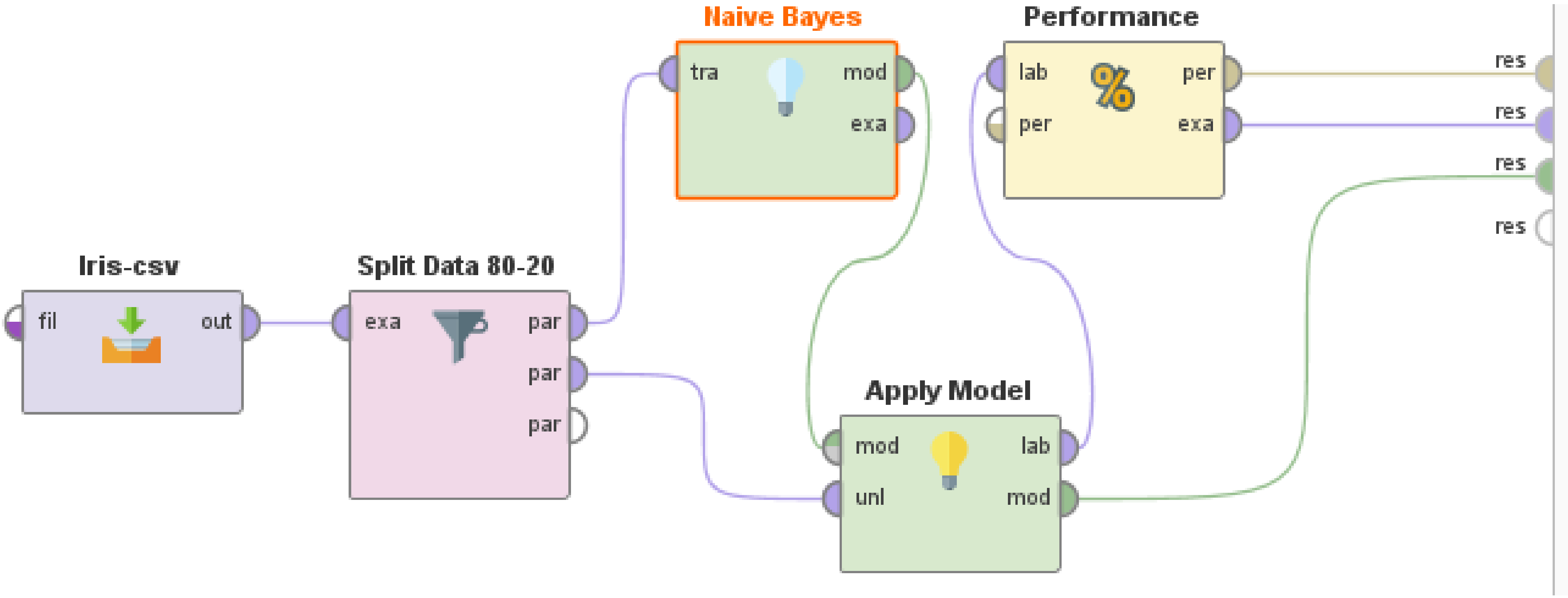
- Se dispone de información de 3 tipos de flores Iris

https://archive.ics.uci.edu/ml/datasets/Iris

Iris.csv

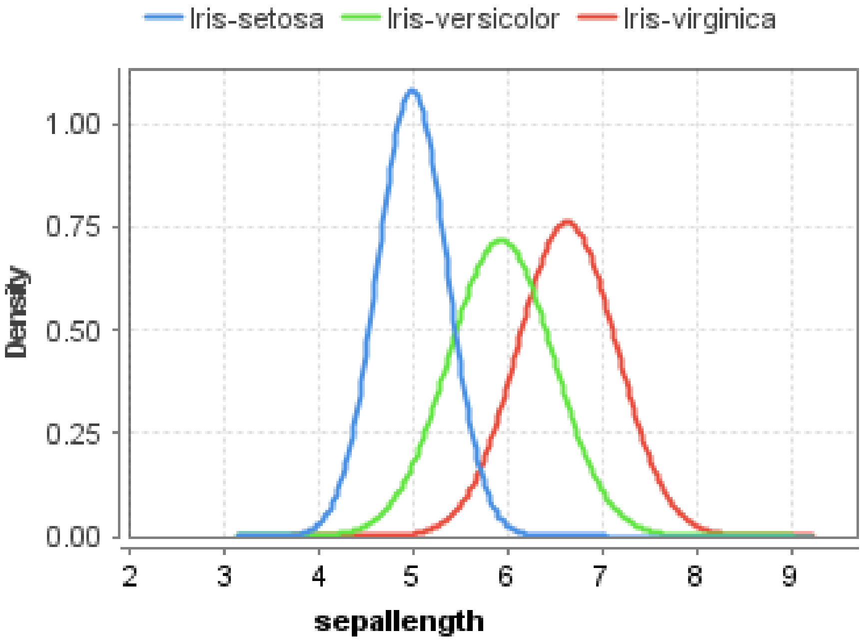
Id	sepallength	sepalwidth	petallength	petalwidth	class
1	5,1	3,5	1,4	0,2	Iris-setosa
2	4,9	3,0	1,4	0,2	Iris-setosa
…	…	…	…	…	…
95	5,6	2,7	4,2	1,3	Iris-versicolor
96	5,7	3,0	4,2	1,2	Iris-versicolor
97	5,7	2,9	4,2	1,3	Iris-versicolor
…	…	…	…	…	…
149	6,2	3,4	5,4	2,3	Iris-virginica
150	5,9	3,0	5,1	1,8	Iris-virginica

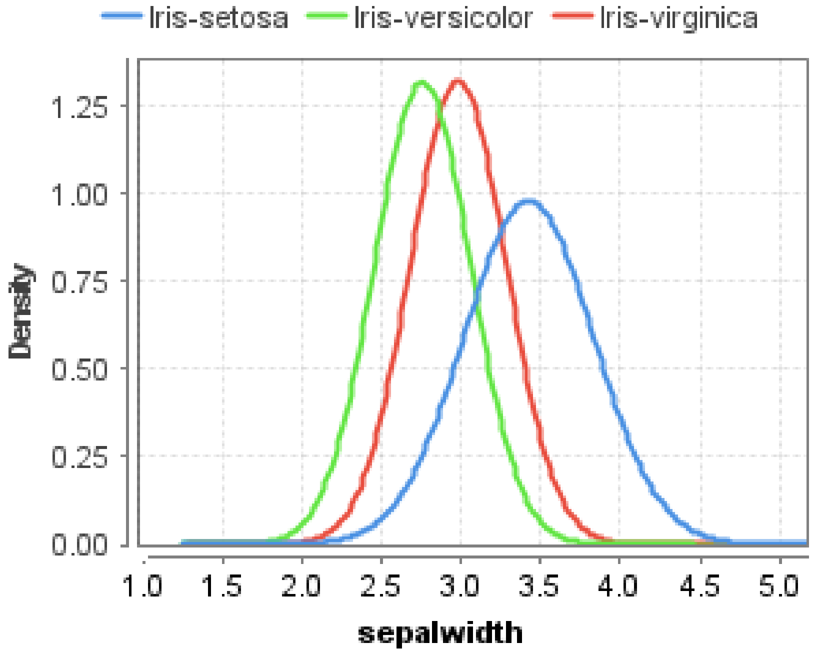
Clasificación de flores de Iris

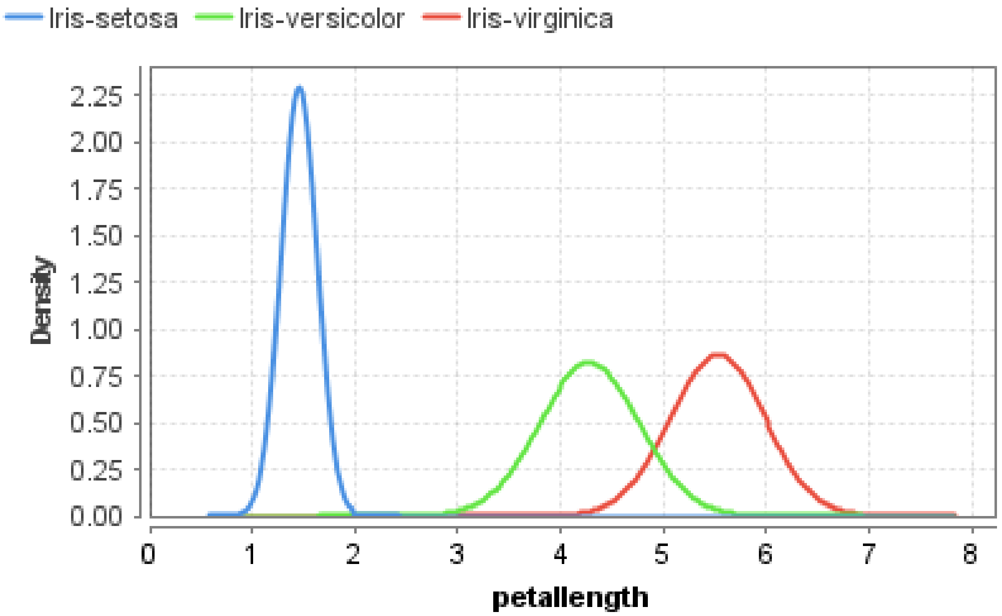


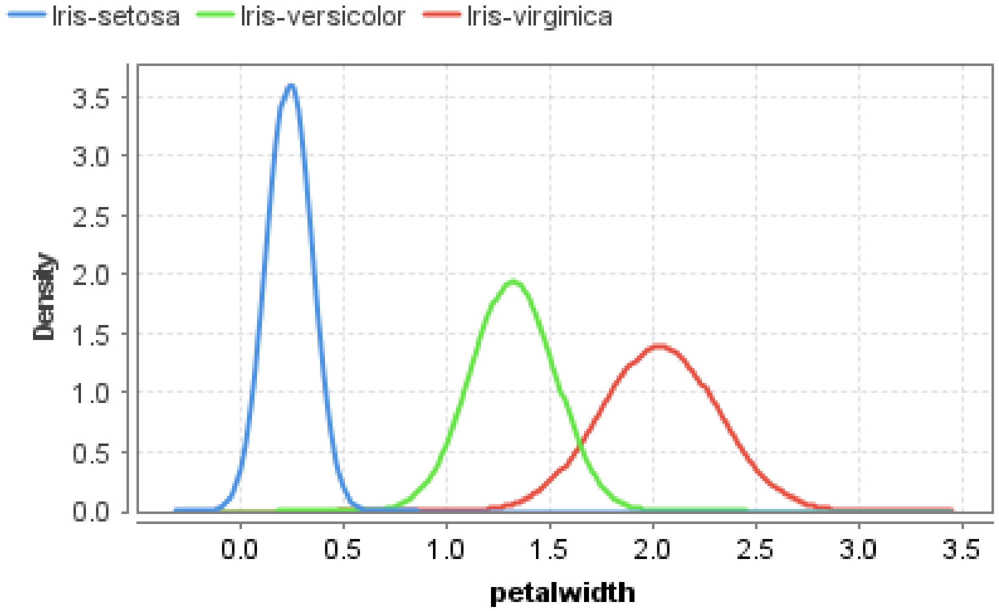
Clasificador Bayesiano

Attribute	Parameter	Iris-setosa	Iris-versicolor	Iris-virginica
sepallength	mean	4.990	5.930	6.630
sepallength	standard deviation	0.369	0.556	0.525
sepalwidth	mean	3.425	2.758	2.982
sepalwidth	standard deviation	0.407	0.303	0.302
petallength	mean	1.452	4.280	5.527
petallength	standard deviation	0.174	0.487	0.464
petalwidth	mean	0.242	1.323	2.032
petalwidth	standard deviation	0.111	0.206	0.286


Iris-setosa
Iris-versicolor
Iris-virginica
Density
0.00
0.25
0.50
0.75
1.00
2
3
4
5
6
7
8
9
sepallength


Iris-setosa
Iris-versicolor
Iris-virginica
Density
0.00
0.25
0.50
0.75
1.00
1.25
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
sepalwidth


Iris-setosa
Iris-versicolor
Iris-virginica
Density
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
0
1
2
3
4
5
6
7
8
petallength


Iris-setosa
Iris-versicolor
Iris-virginica
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
petalwidth

Clasificación de flores de Iris

accuracy: 96.67%

	true Iris-setosa	true Iris-versicolor	true Iris-virginica	class precision
pred. Iris-setosa	10	0	0	100.00%
pred. Iris-versicolor	0	10	1	90.91%
pred. Iris-virginica	0	0	9	100.00%
class recall	100.00%	100.00%	90.00%	