

Naive Bayes

Algoritmo que gracias a un simple supuesto puede ser usado para hacer predicciones

Se base en el teorema de Bayes:

$$P(H|E) = \frac{P(H) \cdot P(E|H)}{P(E)}$$

Pero, ¿qué nos dice este teorema?

Supongamos dos eventos relacionados:

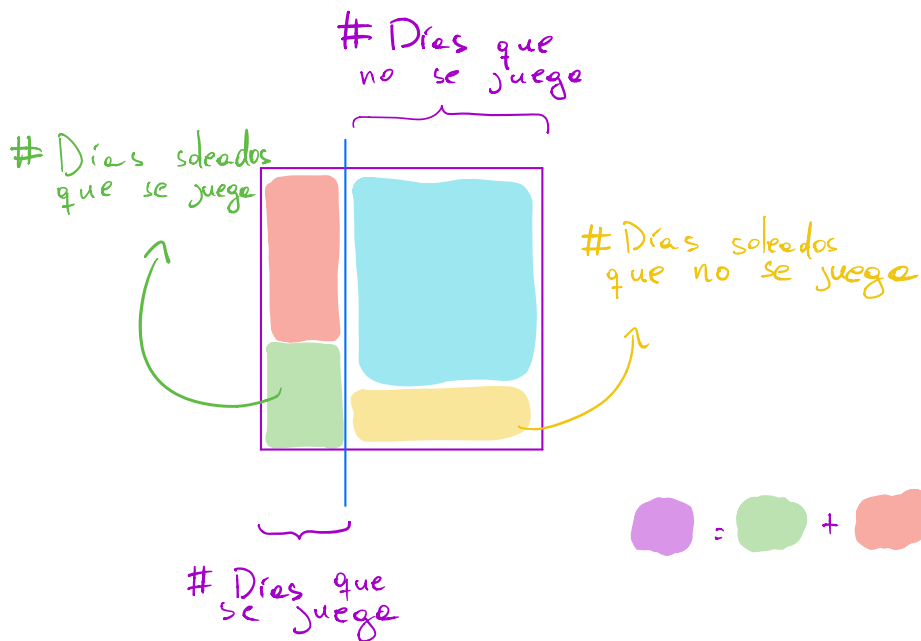
J es el evento jugar el fútbol un día

S es el evento el día está soleado

Entonces:

$$P(J|S) = \frac{P(J) P(S|J)}{P(S)}$$

Intentemos ver esto de forma gráfica:



Queremos $P(J|S)$

$$P(S|J) = \frac{\text{green}}{\text{green} + \text{red}} \quad P(S|\neg J) = \frac{\text{yellow}}{\text{yellow} + \text{blue}}$$

$$P(J|S) = \frac{\text{green}}{\text{green} + \text{yellow}}$$

Pero en términos de probabilidades,
¿Cuanto valen y ?

$$\text{green} = \text{purple} \times P(J) \times P(S|J)$$

$$\text{yellow} = \text{purple} \times P(\neg J) \times P(S|\neg J)$$

$$P(J|S) = \frac{\text{blue} \times P(J) \times P(S|J)}{\text{blue} \times P(J) \times P(S|J) + \text{blue} \times P(\neg J) \times P(S|\neg J)}$$

$$P(J|S) = \frac{P(J) \times P(S|J)}{P(J) \times P(S|J) + P(\neg J) \times P(S|\neg J)}$$

Pero también $\text{green} + \text{yellow} = \text{blue} \times P(S)$

Entonces:

$$P(J|S) = \frac{P(J) P(S|J)}{P(S)}$$

Ahora, ¿cómo usamos esto para clasificar?

Supongamos que queremos predecir si un suceso Y sucederá o no (0, 1) igual a i , dado que sabemos X_1, \dots, X_p .

Esto es:

$$P(Y|X_1, \dots, X_p)$$

Si seguimos la mecánica hasta ahora, tenemos:

$$P(Y=i|X_1, \dots, X_p) = \frac{P(Y=i)P(X_1, \dots, X_p|Y=i)}{P(Y=0)P(X_1, \dots, X_p|Y=0) + P(Y=1)P(X_1, \dots, X_p|Y=1)}$$

Pero aquí usamos la parte "naive":
Vamos a asumir que:

$$P(X_1, \dots, X_p|Y=i) = \prod_{j=1}^p P(X_j|Y=i)$$

$$\underbrace{P(X_1|Y=i) \times \dots \times P(X_p|Y=i)}$$

Veamos un ejemplo:

Queremos predecir si este sábado habrá partido de la liga de computación UAI según estos datos:

Día de la semana	Clima	Partido
Lunes	Sol	No
Martes	Lluvia	No
Miércoles	Lluvia	Si
Jueves	Sol	No
Viernes	Sol	Si
Sábado	Sol	Si
Lunes	Lluvia	Si
Miércoles	Sol	No
Viernes	Sol	No
Sábado	Sol	No
Lunes	Lluvia	Si
Sábado	Lluvia	Si
Martes	Sol	Si
Sábado	Sol	Si
Sábado	Lluvia	No

Queremos $P(J = \text{"Si"} \mid D = \text{"S"}, C = \text{"Sol"})$

=

$$P(J = \text{"Si"}) P(D = \text{"S"} \mid J = \text{"Si"}) P(C = \text{"Sol"} \mid J = \text{"Si"})$$

$$P(J = \text{"Si"}) P(D = \text{"S"} \mid J = \text{"Si"}) P(C = \text{"Sol"} \mid J = \text{"Si"}) + P(J = \text{"No"}) P(D = \text{"S"} \mid J = \text{"No"}) P(C = \text{"Sol"} \mid J = \text{"No"})$$

$$P(J = \text{"Si"}) = \frac{8}{15}$$

$$P(J = \text{"No"}) = \frac{7}{15}$$

$$P(C = \text{"Sol"} | J = \text{"Si"}) = \frac{4}{8}$$

$$P(D = \text{"S"} | J = \text{"Si"}) = \frac{3}{8}$$

$$P(C = \text{"Sol"} | J = \text{"No"}) = \frac{5}{7}$$

$$P(D = \text{"S"} | J = \text{"No"}) = \frac{2}{7}$$

$$P(J = \text{"Si"}) P(D = \text{"S"} | J = \text{"Si"}) P(C = \text{"Sol"} | J = \text{"Si"})$$

$$P(J = \text{"Si"}) P(D = \text{"S"} | J = \text{"Si"}) P(C = \text{"Sol"} | J = \text{"Si"}) + P(J = \text{"No"}) P(D = \text{"S"} | J = \text{"No"}) P(C = \text{"Sol"} | J = \text{"No"})$$

=

$$\frac{8}{15} \cdot \frac{3}{8} \cdot \frac{4}{8}$$

$$\frac{8}{15} \cdot \frac{3}{8} \cdot \frac{4}{8} + \frac{7}{15} \cdot \frac{2}{7} \cdot \frac{5}{7}$$