



# NAÏVE BAYES

---

Abraham Sánchez López  
FCC/BUAP  
Grupo MOVIS

# Introducción

- Naïve Bayes es uno de los algoritmos de clasificación más simples de entender y, por lo tanto, lo desarrollaremos en esta parte del curso.
- Una de las áreas de aplicación clásicas de este modelo es la clasificación de textos basada únicamente en la frecuencia de aparición de las palabras.
- Aunque está basado en la suposición simplista (ingenua) de total independencia de las variables, es un algoritmo versátil que se encuentra naturalmente en la aplicación de diversas industrias, incluso cuando no se considera el supuesto básico!
- En un artículo de 1997, Domingos y Pazzani incluso analizaron los motivos teóricos de esta eficacia contraintuitiva.
- Elegimos explicar su operación mediante una ilustración, en lugar de pasar el tiempo en las bases teóricas relativamente simples de Naïve Bayes.
- Estas bases (el teorema de Bayes y la noción de independencia) se mencionan brevemente al comienzo del tema.
- Finalmente, se presentará una formulación general del modelo Naïve Bayes.

# El teorema de Bayes

- En teoría de probabilidad, el teorema de Bayes enuncia probabilidades condicionales.
- Dados dos eventos  $A$  y  $B$ , permite determinar la probabilidad de que  $A$  conozca a  $B$ , si conocemos las probabilidades de  $A$ , de  $B$  y  $B$  conociendo  $A$ , de acuerdo con la siguiente relación:

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

- El término  $P(A)$  es la probabilidad a priori de  $A$ .
- Es “anterior” en el sentido de que precede a cualquier información sobre  $B$ .
- $P(A)$  también se denomina probabilidad marginal de  $A$ .
- El término  $P(A|B)$  se llama la probabilidad posteriori de  $A$  conociendo  $B$  (o la de  $A$  sub condición de  $B$ ).
- Es “posterior”, en el sentido de que depende directamente de  $B$ .
- El término  $P(B|A)$ , para un  $B$  conocido, se denomina función de verosimilitud de  $A$ .
- Y finalmente, el término  $P(B)$  es la probabilidad marginal o a priori de  $B$ .

# La noción de independencia

- Hablemos de un caso particular de relación entre  $A$  y  $B$ : independencia.
- Se dice que dos eventos  $A$  y  $B$  son independientes cuando la probabilidad de  $A$  no depende de la de  $B$ . En otras palabras:

$$p(A|B) = p(A)$$

- Naïve Bayes se basa en el supuesto de independencia de las variables del problema que se modela.
- Esta es una fuerte suposición: supongamos que estás modelando el comportamiento de tus clientes y has identificado varias variables que discriminan tu problema, como la edad y el salario.
- Es obvio que estas dos variables están correlacionadas y, por lo tanto, que una depende de la otra. Bueno, Naïve Bayes asumirá lo contrario, es decir:

$$p(\text{salario}|\text{edad}) = p(\text{salario})$$

- Ahora veamos cómo funciona el modelo, según esta propiedad y el teorema de Bayes.

# El modelo Naïve Bayes con un ejemplo, I

- Ilustramos el funcionamiento de este modelo con un ejemplo.
- Supongamos que trabajas para un motor de búsqueda.
- Supongamos que necesitas desarrollar un algoritmo de ubicación publicitaria basado en búsquedas de usuarios.
- En particular, debes elegir si deseas enviar un anuncio a aquellos cuya búsqueda contiene la palabra “flor”.
- Tu investigación relacionada con esta palabra puede corresponder al deseo de comprar una flor, pero también a una simple búsqueda de información sobre una flor.
- En general, el usuario especificará su búsqueda escribiendo a veces “flor barata” o “ramo de flores”, por ejemplo, en el primer caso y, a veces, “flor roja” o “¿cuándo regar una flor?” en el segundo.
- Por lo tanto, tu motor de búsqueda tiene una gran base de datos de personas que buscan “flores”.
- También tienes otra información importante: acciones posteriores a búsquedas, como actos de compra reales o al menos clics en sitios de compras.

# El modelo Naïve Bayes con un ejemplo, II

- Esto permite construir una base supervisada, para responder a un problema que afecta a muchos anunciantes: ¿qué búsqueda tiene la mayor probabilidad de ser una señal para anunciar un acto de compra?
- Uno podría, por ejemplo, intentar enviar un anuncio cuando se detecte tal señal.
- Para esto, nos gustaría calcular la probabilidad de compra conociendo las palabras de la búsqueda, es decir  $p(\text{compra}|\text{palabras de la búsqueda})$ , dada la siguiente base de datos ilustrativa en la siguiente tabla.

ID_internauta	Búsqueda que contiene “barato”	Búsqueda que contiene “rojo”	Compra?
1	Si	No	Si
2	No	Si	No
3	No	Si	No
4	No	Si	No
5	Si	No	No
6	Si	Si	Si

# El modelo Naïve Bayes con un ejemplo, III

- Conocemos los estados (SÍ o NO) de las palabras clave “barato” y “rojo” (siempre asociados a la palabra “flor”) para cada usuario, así como su comportamiento de compra.
- Desde esta base, uno puede, por ejemplo, hacer la pregunta: ¿debería enviar un anuncio cuando la búsqueda contiene las palabras clave “barato” y “rojo”?
- Esto es para evaluar  $p(\text{compra} \mid \text{barato}, \text{rojo})$ .
- De acuerdo con el teorema de Bayes, uno puede escribir:

$$p(\text{compra} \mid \text{barato}, \text{rojo}) = \frac{p(\text{barato}, \text{rojo} \mid \text{compra}) * p(\text{compra})}{p(\text{barato}, \text{rojo})} \quad (0)$$

- Ahora es cuando utilizaremos la hipótesis fuerte del clasificador sobre el término en negrita, es decir, la independencia de las variables. Esto significa varias cosas:

$$p(\text{rojo} \mid \text{barato}) = p(\text{rojo}) \quad (1)$$

$$p(\text{barato} \mid \text{rojo}) = p(\text{barato}) \quad (2)$$

$$p(\text{barato}, \text{rojo} \mid \text{compra}) = p(\text{barato} \mid \text{compra}) * p(\text{rojo} \mid \text{compra}) \quad (3)$$

# El modelo Naïve Bayes con un ejemplo, IV

- Es la tercera expresión la que nos interesa. Al usarla para reemplazar el término en negrita de (0), podemos reescribir (0) de la siguiente manera:

$$\begin{aligned} & p(\text{compra} \mid \text{barato}, \text{rojo}) * p(\text{barato}, \text{rojo}) \\ &= p(\text{barato} \mid \text{compra}) * p(\text{rojo} \mid \text{compra}) * p(\text{compra}) \end{aligned}$$

- El término a la derecha de la igualdad nos permitirá evaluar la probabilidad de compra dada la búsqueda del usuario.
- Los valores de los términos  $p(\text{barato} \mid \text{compra})$ ,  $p(\text{rojo} \mid \text{compra})$ ,  $p(\text{compra})$  se pueden encontrar directamente en la base de datos anterior, independientemente del valor de la variable “comprar” (SÍ o NO).
- Por lo tanto, la probabilidad de compra se puede evaluar mediante el cálculo del cociente A siguiente:

$$\begin{aligned} & A(\text{barato}, \text{rojo}) \\ &= \\ & \frac{p(\text{barato} \mid \text{compra} = \text{si}) * p(\text{rojo} \mid \text{compra} = \text{si}) * (\text{compra} = \text{si})}{p(\text{barato} \mid \text{compra} = \text{no}) * p(\text{rojo} \mid \text{compra} = \text{no}) * (\text{compra} = \text{no})} \end{aligned}$$



# El modelo Naïve Bayes con un ejemplo, V

- Una decisión de enviar un anuncio si  $A$  es mayor que un umbral  $\alpha$  (por lo general 1, es decir, hay más compras que las no compras para esta búsqueda).
- Nuestra micro-base de datos proporcionará valores para los siguientes términos:

$$p(\text{barato} \mid \text{compra} = \text{si}) = 2/2 = 1$$

$$p(\text{rojo} \mid \text{compra} = \text{si}) = 1/2 = 0.5$$

$$p(\text{compra} = \text{si}) = 2/6 = 0.33$$

$$p(\text{barato} \mid \text{compra} = \text{no}) = 1/4 = 0.25$$

$$p(\text{rojo} \mid \text{compra} = \text{no}) = 3/4 = 0.75$$

$$p(\text{compra} = \text{no}) = 4/6 = 0.66$$

- Podemos calcular  $A$  para la búsqueda de “flor roja barata”:

$$A(\text{barato}, \text{rojo}) = \frac{1 * 0.5 * 0.33}{0.25 * 0.75 * 0.66} = 1.33$$

- $A(\text{barato}, \text{rojo}) > 1$ , por lo que es apropiado para enviar un anuncio. Según el mismo principio, podemos calcular  $A$  para otras búsquedas:

# El modelo Naïve Bayes con un ejemplo, VI

$$A(\text{barato}) = \frac{1 * 0.33}{0.25 * 0.66} = 2$$

$$A(\text{rojo}) = \frac{0.5 * 0.33}{0.75 * 0.66} = 0.33$$

- Por supuesto, existe una alta proporción de “flor cara”, lo que significa que la presencia del término “barato” es muy discriminante en comparación con nuestra pregunta de búsqueda de señales de compra ( $2 \gg 1$ ).
- En cuanto a este, el término “rojo” claramente no anuncia el deseo de comprar, sino que refleja una búsqueda simple de información ( $0.33 \ll 1$ ).
- En el primer caso presentado, “flor roja barata”, observamos que la presencia de “barato” lo “lleva” sobre la presencia de “rojo” y es por eso que también ofreceremos publicidad.
- Puedes terminar de convencerte de todo esto buscando en Google estas tres expresiones.

# El marco general

- Más generalmente, y de acuerdo con una formulación un poco más rigurosa, Naïve Bayes se puede definir de la siguiente manera.
- Sean  $N$  variables  $(X_1, X_2, \dots, X_n)$  que toman los valores  $x_{ij}$ , sobre la base de lo cual tratamos de tomar una decisión binaria  $d \in \{0, 1\}$ .

- Se calcula la relación de probabilidad  $A$  definida por:

$$A = \prod_{i=1}^n \frac{p(x_i|d=1)}{p(x_i|d=0)} * \frac{p(d=1)}{p(d=0)}$$

- La toma de decisión es entonces:

$$d = \begin{cases} 1 & \text{si } A > \alpha \\ 0 & \text{si } A \leq \alpha \end{cases}$$

- Ten en cuenta que Naïve Bayes se convierte en un clasificador lineal cuando se expresa en un espacio logarítmico ( $\log \prod_i x_i = \sum_i \log x_i$ ), pero luego hablaremos de la linealidad de los modelos.

# Resumen

- Aprendizaje supervisado - clasificación
- Naïve Bayes es uno de los clasificadores más simples.
- Se basa en una fuerte hipótesis: la independencia de las variables.