

Módulo 5 – Aprendizaje de Máquina Supervisado

Modelos Bayesianos

Especialización en Ciencia de Datos

Modelos Bayesianos



Thomas Bayes

- Thomas Bayes (1701-1761) fue un matemático y ministro presbiteriano inglés, mejor conocido por su desarrollo de la teoría de la probabilidad bayesiana. Estudió en la Universidad de Edimburgo y más tarde se convirtió en miembro de la Royal Society. La contribución más famosa de Bayes a las matemáticas es su trabajo publicado póstumamente "Ensayo hacia la solución de un problema en la doctrina de las posibilidades", en el que introdujo una fórmula para calcular la probabilidad de un evento en función del conocimiento o la experiencia previa.
- La fórmula, ahora conocida como teorema de Bayes, es un concepto fundamental en la teoría de la probabilidad y la estadística. Proporciona una forma de actualizar nuestras creencias sobre la probabilidad de un evento en función de nueva evidencia o información. El teorema de Bayes tiene aplicaciones en muchas áreas, incluyendo la ciencia, la ingeniería, las finanzas y la medicina.

Aunque el teorema de Bayes no fue ampliamente reconocido durante su vida, desde entonces se ha convertido en una herramienta esencial para la toma de decisiones y la predicción. Hoy en día, el teorema de Bayes se utiliza ampliamente en el aprendizaje automático y la inteligencia artificial, y sigue siendo un área activa de investigación.



Thomas Bayes

- El teorema de Bayes es un procedimiento que nos permite expresar la probabilidad condicional de un evento aleatorio A dado B, en términos de la distribución de probabilidad del evento B dado A y la distribución de probabilidad de solo A.
- Este teorema es de mucha utilidad, ya que gracias a él podemos relacionar la probabilidad de que un evento A ocurra sabiendo que ocurrió B, con la probabilidad de que ocurra lo contrario, es decir, que ocurra B dado A.
- Bayes se ocupó de este teorema en un trabajo titulado “An Essay towards solving a Problem in the Doctrine of Chances” (Un ensayo para resolver un problema en la doctrina de las posibilidades), publicado en 1763, y sobre el cual se han desarrollado grandes estudios con aplicaciones en diversas áreas de conocimiento.

THE DOCTRINE OF CHANCES:

OR,
A Method of Calculating the Probability
of Events in Play.



By *A. De Moivre*. F. R. S.

L O N D O N:

Printed by W. Freeman, for the Author. MDCCLXIII.

Teorema de Bayes

El teorema de Bayes es una manera de calcular una probabilidad cuando conocemos otras ciertas probabilidades. La fórmula es:

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

Lo cual nos dice:

Con qué frecuencia sucede A dado que sucede B, escrito como $P(A|B)$.

Cuando sabemos:

- Con qué frecuencia sucede B dado que sucede A, escrito como $P(B|A)$.
- Qué tan probable es A por sí mismo, lo cual se escribe como $P(A)$.
- Qué tan probable es B por sí mismo, lo cual se escribe $P(B)$.



Teorema de Bayes

Veamos un ejemplo:

Supongamos que:

- $P(F)$ es la probabilidad de tener fuego, es decir, con qué frecuencia hay fuego.
- $P(H)$ es la probabilidad de tener humo, es decir, con qué frecuencia vemos humo.

Por lo tanto,

- $P(F|H)$ corresponde a la probabilidad que haya fuego cuando hay humo. Es decir, con qué frecuencia veo fuego cuando hay humo.
- $P(H|F)$ corresponde a la probabilidad que haya humo cuando hay fuego. Es decir, con qué frecuencia veo humo cuando hay fuego.

Entonces, la fórmula nos dice las cosas “en un sentido”, por ejemplo $P(F|H)$, cuando conocemos el “sentido contrario”, $P(H|F)$.



Teorema de Bayes

Por ejemplo:

- La probabilidad de un incendio peligroso es baja (1%).
 - La probabilidad de ver humo es bastante común (10%), debido a las estufas, la leña, los asados, etc.
 - El 90% de los incendios peligrosos producen humo.
- Aplicando el teorema de bayes podemos descubrir la probabilidad de fuego peligroso cuando hay humo:

$$P(F|H) = \frac{P(F)P(H|F)}{P(H)}$$

$$P(F|H) = \frac{1\% \cdot 90\%}{10\%}$$

$$P(F|H) = 9\%$$



Teorema de Bayes

Otro ejemplo:

- Sabemos que el 50% de los días lluviosos comienzan nublados.
 - Por otro lado, las mañanas nubladas son comunes, aproximadamente el 40% de los días amanece nublado.
 - Estamos en un mes seco, sólo 3 de 30 días llueve, es decir un 10%.
- Esta mañana amaneció nublada, ¿cuál es la probabilidad de lluvia durante el día?

LL: Lluvia.

NB: Nublado.

$$P(LL|NB) = \frac{P(LL)P(NB|LL)}{P(NB)}$$

$$P(LL|NB) = \frac{10\% \cdot 50\%}{40\%}$$

$$P(LL|NB) = 12.5\%$$



Teorema de Bayes

Hemos visto ejemplos en donde sólo hay dos posibilidades (llueve, no llueve), pero ¿qué pasa si hay más de dos posibilidades?.

Veamos el siguiente ejemplo:

Amanda, Boris y Carlos participan en un concurso de arte:

- Amanda participa con 15 pinturas y el 4% de sus obras han ganado el primer premio.
- Boris participa con 5 pinturas y el 6% de sus obras han ganado en primer premio.
- Cecilia participa con 10 pinturas y el 3% de sus obras ha ganado el primer premio.

➤ ¿Cuál es la probabilidad que Amanda gane el primer lugar?

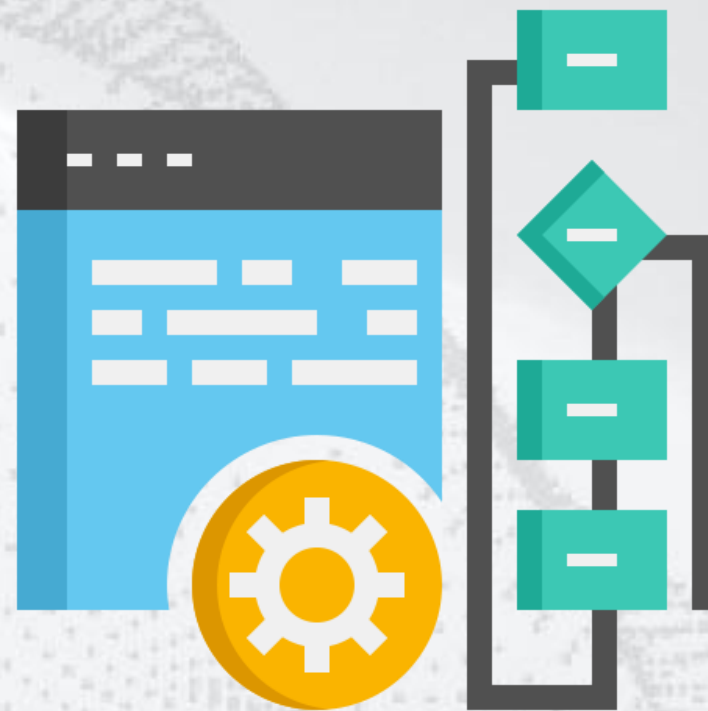
A: Amanda; **B:** Boris; **C:** Cecilia; **L:** Primer Lugar.

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

$$P(A|L) = \frac{\left(\frac{15}{30}\right) \cdot 4\%}{\left(\frac{15}{30}\right) \cdot 4\% + \left(\frac{5}{30}\right) \cdot 6\% + \left(\frac{10}{30}\right) \cdot 3\%}$$

$$P(A|L) = 50\%$$

Naive Bayes



¿Qué es un modelo Naive Bayes?

- En un sentido amplio, los modelos de Naive Bayes son una clase especial de algoritmos de clasificación de Machine Learning, tal y como nos referiremos de ahora en adelante. Se basan en una técnica de clasificación estadística llamada “teorema de Bayes”.
- Estos modelos son llamados algoritmos “Naive”, o “Inocentes”, “Ingenuo” en español. En ellos se asume que **las variables predictoras son independientes entre sí**. En otras palabras, que la presencia de una cierta característica en un conjunto de datos no está en absoluto relacionada con la presencia de cualquier otra característica.
- Proporcionan una manera fácil de construir modelos con un comportamiento muy bueno debido a su simplicidad.



Algoritmo Supervisado Naive Bayes

Si tomamos un umbral de error de A continuación, se listan los pasos para utilizar el algoritmo de Naive Bayes en un problema de clasificación:

Paso 1



Paso 2



Paso 3



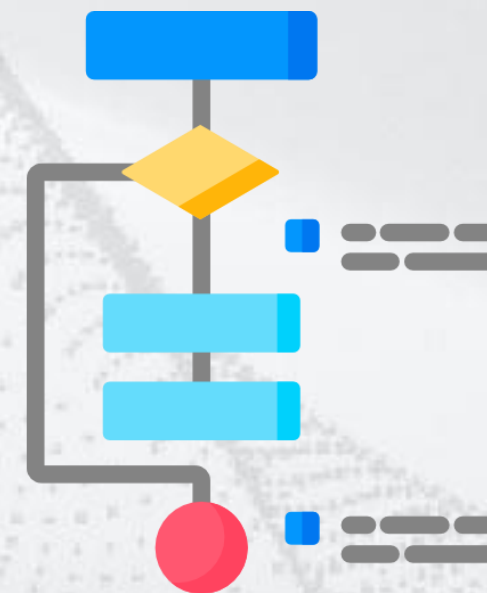
Paso 4

Convertir el conjunto de datos en una tabla de frecuencia.

Crear una tabla de probabilidad calculando las correspondientes a que ocurran los diversos eventos.

La ecuación Naive Bayes se usa para calcular la probabilidad posterior de cada clase.

La clase con la probabilidad posterior más alta es el resultado de la predicción.



Corrección de Laplace

- La corrección de Laplace, también conocida como suavizado de Laplace, o adición de Laplace, es una técnica utilizada en el algoritmo Naive Bayes para **evitar la probabilidad nula o cero** en casos en los que una característica de una instancia no aparece en el conjunto de entrenamiento.
- En el algoritmo Naive Bayes, se calcula la probabilidad condicional de una característica dado una clase determinada. Si una característica no aparece en el conjunto de entrenamiento para una clase en particular, la probabilidad condicional de esa característica se convierte en cero. Esto puede ser un problema en casos en los que hay muchas características y el conjunto de entrenamiento es pequeño.
- La corrección de Laplace resuelve este problema al agregar una constante a cada recuento de características en el cálculo de la probabilidad condicional. Esta constante suele ser 1, lo que se conoce como corrección de Laplace unitaria. Al agregar una constante, se asegura que cada característica tenga una probabilidad no nula, incluso si no aparece en el conjunto de entrenamiento.



Naive Bayes Gaussiano

- Naive Bayes Gaussiano es una variante del algoritmo Naive Bayes que **se utiliza para clasificar datos continuos o numéricos**. En contraste con el Naive Bayes clásico, que asume que las características son categóricas o discretas, el Naive Bayes Gaussiano asume que las características siguen una distribución normal o Gaussiana.
- En el Naive Bayes Gaussiano, se calculan las probabilidades de pertenencia a cada clase para una instancia determinada utilizando la fórmula de Bayes y **asumiendo que las características son independientes entre sí y siguen una distribución normal**. Para cada clase, se calcula la media y la desviación estándar de cada característica numérica en el conjunto de entrenamiento. Estos parámetros se utilizan para calcular la probabilidad condicional de cada característica para cada clase, suponiendo que sigue una distribución normal.



Naive Bayes Gaussiano

- El Naive Bayes Gaussiano es especialmente útil cuando se tienen **conjuntos de datos con muchas características numéricas y una distribución normal**. También puede funcionar bien incluso cuando las características no siguen estrictamente una distribución normal, siempre y cuando la desviación estándar no sea demasiado grande y los valores atípicos sean manejables.
- Ahora que suponemos que la probabilidad de las características es gaussiana, entonces, la probabilidad condicional viene dada por:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$





Ventajas

1. Es un algoritmo rápido y eficiente en términos de tiempo de entrenamiento y predicción, lo que lo hace adecuado para grandes conjuntos de datos.
2. Es fácil de implementar y requiere muy pocos parámetros para ajustar.
3. Funciona bien incluso con un número limitado de datos de entrenamiento.
4. Es robusto ante la presencia de características irrelevantes en los datos, lo que significa que no se ve afectado por la inclusión de características que no contribuyen a la clasificación.
5. Es un clasificador probabilístico, lo que significa que proporciona una estimación de la probabilidad de que una instancia pertenezca a una categoría determinada.



Desventajas

1. Aunque son unos clasificadores bastante buenos, los algoritmos Naive Bayes son conocidos por ser pobres estimadores. Por ello, no se deben tomar muy en serio las probabilidades que se obtienen.
2. El algoritmo Naive Bayes asume que todas las características son independientes entre sí, lo que a veces puede ser una simplificación demasiado extrema de la realidad.
3. No es muy preciso en comparación con otros algoritmos más complejos como los árboles de decisión o las redes neuronales.
4. Si hay características importantes que faltan en los datos de entrenamiento, Naive Bayes puede tener problemas para clasificar correctamente las nuevas instancias que contienen esas características faltantes.

The background of the slide features a light gray, wavy, dotted pattern that resembles a stylized landscape or a textured surface. A solid purple rectangular banner is positioned horizontally across the middle of the image, containing the text "Dudas y consultas" in white.

Dudas y consultas

Fin Presentación