

MACHINE LEARNING & BIG DATA

Supervised Learning: Clasificador Bayesiano

Algoritmos Supervisados

JOSÉ NELSON ZEPEDA DOÑO

Cluster de Estudio: Advanced Analytics

Este material es el resumen de muchos autores que por medio de sus libros y documentos nos ofrecen fuentes riquísimas de conocimiento sobre los temas de Big Data y Machine Learning.

Algunas citas, figuras y tablas pueden ser encontradas de forma textual tal como lo indica el autor en su material original.

Nelson Zepeda

MIP • V 1.0

San Salvador El Salvador

Phone 503 79074137 • @nelsonzepeda733

Tabla de Contenido

Teorema de Bayes	1
Aplicaciones del Teorema de Bayes	2
Debilidades del Teorema de Bayes	3
Aplicaciones del Teorema de Bayes	3
Condiciones del Teorema de Bayes	3
Ventajas del Teorema de Bayes	4
Importancia del Teorema de Bayes	4
Ejemplos de cálculos con el Teorema de Bayes.....	5
Clasificador Bayesiano	7
Cómo funciona el algoritmo.	8
Pros del algoritmo.	9
Cons del algoritmo.	10
Aplicaciones del algoritmo.	10
Bibliografía	11

Teorema de Bayes

Proposición planteada por el matemático inglés Thomas Bayes (1702-1761)¹ en 1763.

El teorema de Bayes, en la teoría de la probabilidad, es una proposición planteada por el matemático inglés Thomas Bayes (1702-1761)¹ en 1763,² que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A.

En términos más generales y menos matemáticos, el teorema de Bayes es de enorme relevancia puesto que vincula la probabilidad de A dado B con la probabilidad de B dado A. Es decir, por ejemplo, que sabiendo la probabilidad de tener un dolor de cabeza dado que se tiene gripe, se podría saber (si se tiene algún dato más), la probabilidad de tener gripe si se tiene un dolor de cabeza. Muestra este sencillo ejemplo la alta relevancia del teorema en cuestión para la ciencia en todas sus ramas, puesto que tiene vinculación íntima con la comprensión de la probabilidad de aspectos causales dados los efectos observados¹.

El teorema de Bayes es utilizado para calcular la probabilidad de un suceso, teniendo información del estado previo o bien conocer de antemano las condiciones sobre ese suceso, es decir, mide la probabilidad de que una variable aleatoria pueda tomar un valor particular dado el valor de otra variable aleatoria.

Se puede calcular la probabilidad de un suceso A, sabiendo además que ese A cumple cierta característica que condiciona su probabilidad. El teorema de Bayes entiende la probabilidad de forma inversa al teorema de la probabilidad total. El teorema de la probabilidad total hace inferencia sobre un suceso B, a partir de los resultados de los sucesos A. Por su parte, Bayes calcula la probabilidad de A condicionado a B.

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

donde:

- $P(A_i)$ son las probabilidades a priori,
- $P(B|A_i)$ es la probabilidad de B en la hipótesis A_i ,
- $P(A_i|B)$ son las probabilidades a posteriori.

¹ https://es.wikipedia.org/wiki/Teorema_de_Bayes

Matemáticamente el teorema de Bayes es igual al cociente del producto de la probabilidad “B” dados (A_i), $P(B/A_i)$ (siendo B el suceso conocido y “ A_i ” los sucesos condicionados) por la probabilidad $P(A_i)$ entre la sumatoria de cada probabilidad que contenga el suceso conocido por cada suceso conocido.

Condiciones para aplicar el teorema de Bayes²

Los sucesos “ A_i ” deben ser mutuamente excluyentes, es decir solo puede suceder uno de ellos.

La unión de sus posibilidades es el total, es decir la unidad es decir debe ser un sistema completo. Y cada una debe ser distinta de cero.

Está establecido un caso “B” del cual se conocen todas las probabilidades.

Se conocen todas las probabilidades condicionales $P(B/A_i)$.

Aplicaciones del Teorema de Bayes

El teorema de Bayes es válido en todas las aplicaciones de la teoría de la probabilidad. Sin embargo, hay una controversia sobre el tipo de probabilidades que emplea. En esencia, los seguidores de la estadística tradicional sólo admiten probabilidades basadas en experimentos repetibles y que tengan una confirmación empírica mientras que los llamados estadísticos bayesianos permiten probabilidades subjetivas. El teorema puede servir entonces para indicar cómo debemos modificar nuestras probabilidades subjetivas cuando recibimos información adicional de un experimento. La estadística bayesiana está demostrando su utilidad en ciertas estimaciones basadas en el conocimiento subjetivo a priori y el hecho de permitir revisar esas estimaciones en función de la evidencia empírica es lo que está abriendo nuevas formas de hacer conocimiento. Una aplicación de esto son los clasificadores bayesianos que son frecuentemente usados en implementaciones de filtros de correo basura o spam, que se adaptan con el uso. Otra aplicación se encuentra en la fusión de datos, combinando información expresada en términos de densidad de probabilidad proveniente de distintos sensores.

Aplicaciones puntuales:

- El diagnóstico de cáncer.
- Evaluación de probabilidades durante el desarrollo de un juego de bridge por Dan F. Waugh y Frederick V. Waugh.
- Probabilidades a priori y a posteriori.
- Un uso controvertido en la Ley de sucesión de Laplace.
- En el testeo de hipótesis en Ciencia Política cuando se usa metodología process tracing.

² <https://www.webyempresas.com/teorema-de-bayes/>

En el campo estadístico el teorema de Bayes permitió la resolución de problemas de múltiples probabilidades, su importancia radica en la aplicación de la misma, pues es fundamental en cualquier ciencia, ya que permite demostrar la relación intrínseca con la comprensión de las probabilidades de los sucesos causados una vez establecidos los efectos acontecidos.

La probabilidad Bayesiana permite convertir una probabilidad subjetiva en una real cuando esta se va modificando en base a las nuevas informaciones.

Las evidencias empíricas que según los estadistas actúan como base para la aplicación de este teorema tienen aplicaciones puntuales en las distintas ramas de la medicina, desde el diagnóstico de cáncer, hasta para la prevención de la diabetes, también tiene usos menos sofisticado como el evaluar las posibilidades en un juego de barajas³.

Debilidades del Teorema de Bayes

Los estadistas han cuestionado el teorema basándose en las limitaciones de su aplicación, ya que es válido únicamente cuando se cumplen sucesos disjuntos y exhaustivos.

De igual manera los especialistas en estadística tradicional ratifican que solo pueden ser admisibles las estadísticas basadas en experimentos repetibles y de comprobación empírica, debido a que las probabilidades estadísticas Bayesianas admiten condiciones relativas.

Aplicaciones del Teorema de Bayes

El teorema de Bayes sirve para calcular las posibilidades de un suceso que está dado o no por otro suceso anterior, lo cual consiente evaluar de qué manera se transforman las probabilidades subjetivas, mientras más información nueva se posee de un hecho.

Además de ser aplicable a modelos basados en el conocimiento subjetivo y la evidencia empírica. Se aplica también a modelos que se utilizan, por ejemplo, en la fusión de datos de un sistema.

Así mismo, es considerado como un excelente modelo o método para evaluar nueva información y revisar estimaciones anteriores sustentadas en datos limitados, para conocer entonces si se encuentran en un estado u otro, si es aplicado de manera idónea entonces se hace eficaz la reunión de datos para tomar mejores decisiones.

Condiciones del Teorema de Bayes

- Los sucesos “ A_i ” deben ser mutuamente excluyentes, es decir solo puede suceder uno de ellos.

³ <https://www.webyempresas.com/teorema-de-bayes/>

- La unión de sus posibilidades es el total, es decir la unidad es decir debe ser un sistema completo. Y cada una debe ser distinta de cero.
- Está establecido un caso “B” del cual se conocen todas las probabilidades.
- Se conocen todas las probabilidades condicionales $P(B/A_i)$.

Ventajas del Teorema de Bayes

- Se puede enfocar de manera tal de que se obtengan beneficios en algunos campos.
- Es posible el análisis continuo de la información, aunque si la variabilidad entre datos es elevada es necesario algún método que permita llegar a soluciones confiables.
- Meta-Análisis: buscar acumular información variada para llegar a una apreciación exacta de un problema
- Evaluación de estudios a pequeña escala con la información de otros, debido a que el desarrollo de estos a escala global no siempre es posible, y a nivel muestra no cuenta con una total veracidad, el enfoque bayesiano permite ratificar y refutar.
- Estudios de decisión.

Importancia del Teorema de Bayes

En el campo estadístico el teorema de Bayes permitió la resolución de problemas de múltiples probabilidades, su importancia radica en la aplicación de la misma, pues es fundamental en cualquier ciencia, ya que permite demostrar la relación intrínseca con la comprensión de las probabilidades de los sucesos causados una vez establecidos los efectos acontecidos. La probabilidad Bayesiana permite convertir una probabilidad subjetiva en una real cuando esta se va modificando en base a las nuevas informaciones.

Las evidencias empíricas que según los estadistas actúan como base para la aplicación de este teorema tienen aplicaciones puntuales en las distintas ramas de la medicina, desde el diagnóstico de cáncer, hasta para la prevención de la diabetes, también tiene usos menos sofisticado como el evaluar las posibilidades en un juego de barajas. Recapitulando, este teorema sirve para evaluar sucesos a priori y a posteriori, teniendo en cuenta hechos que pueden ser subjetivos o no y en base a las posibilidades que desencadenen estos hechos, obtener un dato que como conocimiento permitirá o no establecer un plan de acción.

Ejemplos de cálculos con el Teorema de Bayes

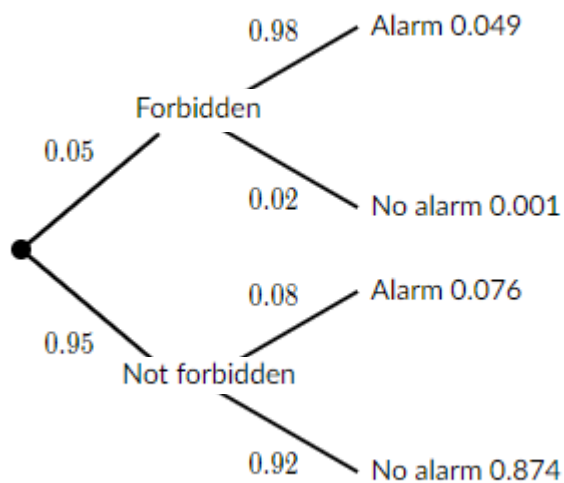
Equipaje en el aeropuerto⁴:

En el aeropuerto existen métodos para detectar artículos prohibidos, una alarma se activara cuando un artículo prohibido aparezca en alguna de las maletas.

- Estadísticamente, el 5% de las maletas contienen artículos prohibidos.
- Dado que la alarma no es perfecta, si una maleta contiene artículos prohibidos, existe un 98% de probabilidades de que la alarma se active.
- Si una maleta no contiene artículos prohibidos, existe un 8% de probabilidades que la alarma se dispare.

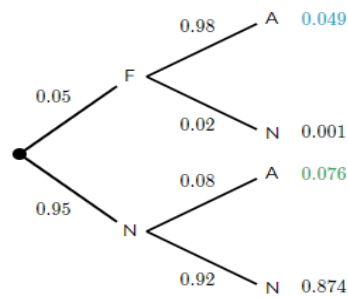
¿Dado un grupo aleatorio de maletas que activaron la alarma, cual es la probabilidad que estas contengan un artículo prohibido?

Primero armemos el árbol de decisiones:



⁴ <https://www.khanacademy.org/math/ap-statistics/probability-ap/stats-conditional-probability/a/tree-diagrams-conditional-probability>

La Probabilidad de que la alarma se active se puede dar en 2 escenarios:



$$P(A) = P(F \cap A) + P(N \cap A)$$

$$= 0.049 + 0.076$$

$$= 0.125$$

Volviendo a la pregunta original, dado un grupo de maletas que activaron la alarma, ¿cuál es la probabilidad de que la maleta contenga un objeto prohibido?

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

$$= \frac{0.049}{0.049 + 0.076}$$

$$= \frac{0.049}{0.125}$$

$$= 0.392$$

Clasificador Bayesiano

Naive Bayes utiliza datos históricos para encontrar asociaciones y relaciones y hacer predicciones.

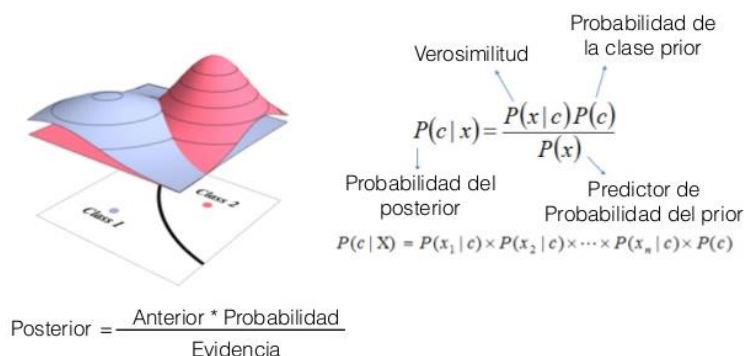
Naive Bayes es una técnica de clasificación y predicción que construye modelos que predicen la probabilidad de posibles resultados.

Recibe el nombre de Clasificador Ingenuo pues asume que la presencia o ausencia de una característica particular no está relacionada con la presencia o ausencia de cualquier otra característica.

Por ejemplo, una fruta puede ser considerada como una manzana si es roja, redonda y alrededor de 7cm de diámetro, esto se debe a que un clasificador de Bayes ingenuo, considera que cada una de estas características contribuye de manera independiente a la probabilidad de que esta fruta sea una manzana, independientemente de la presencia o ausencia de las otras características.

En un entorno de aprendizaje supervisado, los clasificadores de bayes ingenuo, se pueden entrenar de manera muy eficiente pues solo requiere una cantidad pequeña de datos para el entrenamiento para estimar los parámetros (las medias y las varianzas de las variables) necesarias para la clasificación.

Dado que las variables independientes se asumen, solo es mandatorio determinar las varianzas de las variables de cada clase y no toda la matriz de covarianza⁵.



⁵ <https://prezi.com/16ddv65jhzzi/algorithm-naive-bayes/>

El clasificador Bayesiano predice la pertenencia de una observación a una determinada “clase” en función del cálculo de sus probabilidades, es decir la clase que tenga la probabilidad más alta será la respuesta entregada por el clasificador, matemáticamente hablando, esto es conocido como MAP o Max A Posteriori.

Cómo funciona el algoritmo.

Se explicara el funcionamiento mediante un ejemplo:

Supongamos que tenemos un dataset de entrenamiento que contiene algunos datos del clima y que la variable objetivo se denomina “Play”, nuestro objetivo es determinar si se jugara o no se jugara un partido de golf en función de las condiciones climatológicas.

A continuación el contenido de nuestro dataset.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

El primer paso es convertir el dataset en una tabla de frecuencias (Probabilidad conjunta)

Weather	No	Yes
Overcast	0	4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Supongamos que se necesita contestar si es correcto afirmar que se jugara el partido dado que esta soleado.

Primero se calculara la probabilidad que el día este soleado:

$$P(\text{Sunny}) = 5/14 = 0.36,$$

Segundo, calcular la probabilidad de que si se jugara dado que esta soleado:

$$(P(\text{Sunny} | \text{Yes})) = 3/9 = 0.33,$$

Tercero se necesita la probabilidad de que en general si se jugara:

$$P(\text{Yes}) = 9/14 = 0.64$$

Y por último calcular la probabilidad condicional:

$$P(\text{Yes} | \text{Sunny}) = 0.33 * 0.64 / 0.36 = 0.60$$

Esto muestra que la probabilidad es alta y se puede contestar a la pregunta inicial del problema.

Los mismos cálculos se realizan cuando existen diferentes clases y mayor cantidad de atributos.

Pros del algoritmo.

1. Permite realizar predicciones de forma rápida y con complejidad mínima.
2. Se puede utilizar en problemas con una sola clase o múltiples clases.
3. Cuando el supuesto de independencia es verdadero, el clasificador bayesiano funciona muy bien y de forma muy eficiente en comparación a otros modelos.

4. Se requiere menos data para entrenamiento
5. Las variables categóricas funcionan sin mayor problema.
6. Las variables numéricas asumen una distribución normal.

Cons del algoritmo.

1. Si existe una variable categórica y alguna de sus clases no se encuentra dentro de las observaciones del dataset, el modelo le asigna una probabilidad de valor cero lo que implica que esa clase formaría parte de una predicción. Para solventar este problema se utiliza la transformada de Laplace.
2. El clasificador bayesiano asume que las variables predictoras son independientes, es la vida real, es muy complicado poder garantizar que las variables sean independientes entre si.

Aplicaciones del algoritmo.

1. Predicciones en tiempo real debido a su velocidad
2. Predicciones multiclases
3. Clasificación de textos, análisis de sentimientos, análisis de Spam.
4. Modelos de recomendación

Bibliografía

- Machine Learning for Beginners
By Ken Richards, 2017
- Modelado mediante RF de las emisiones de autobuses urbanos en función de los ciclos cinemáticos
By Victor Pita González-Campos , 2017
- R Data Analysis Cookbook
by Kuntal Ganguly, 2017