Práctica 6: Algoritmo de Naive Bayes

Profesor: Lauro Reyes Cocoletzi

Edkir Nava y Diego Castro
{enavam2001, dcastroe2100} @alumno.ipnx.mx

UPIIT: Unidad Profesional Interdisciplinaria en Ingeniería Campus Tlaxcala Instituto Politécnico Nacional, Tlaxcala, Tlaxcala, México 9000

Ingeniera en Inteligencia Artificial

15 de noviembre 2023

***Resumen*— El teorema de Bayes, fundamental en estadística y aprendizaje automático, se utiliza en clasificación mediante algoritmos como Naive Bayes. La aplicación de Naive Bayes implica calcular las probabilidades condicionales de las características dadas las clases para realizar predicciones. El método A Priori, esencial en este contexto, evalúa la probabilidad de las clases sin considerar las características específicas. Por otro lado, la implementación multivariable implica la consideración simultánea de múltiples características en el proceso de clasificación. Al integrar estas probabilidades condicionales, Naive Bayes logra una clasificación eficiente y precisa, siendo particularmente útil en conjuntos de datos con múltiples variables y en situaciones donde las características son independientes dadas las clases.**

***Palabras clave* — Teorema de Bayes, multivariable, probabilidad, frecuencia, verosimilitud**

I. Marco Teorico

A. Teorema de Bayes

El cálculo de una probabilidad posterior $P(A|B)$ a partir de probabilidades previas dadas $P(A_i)$ y probabilidades condicionales $P(A|B)$ ocupa una oposición central en la probabilidad elemental. La regla general de dichos cálculos, los que en realidad son una aplicación simple de la regla de multiplicación, se remonta al reverendo Thomas Bayes quien vivió en el siglo XVIII.

Para formularla primero se requiere otro resultado previo. Sean $A_1, \dots, A_k$ un conjunto de eventos mutuamente excluyentes y exhaustivos con probabilidades *previas* $P(A_i)(i = 1, \dots, k)$. Entonces para cualquier otro evento B para el cual $P(B) > 0$, la probabilidad *posterior* de A_j dado que ha ocurrido B es:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \qquad (1)$$

B. Teorema de Bayes Multivariable

El Teorema de Bayes puede extenderse al caso multivariable cuando se tienen múltiples variables aleatorias en juego. En lugar de lidiar con eventos simples, se trabajará con vectores de eventos. La forma general del Teorema de Bayes para el caso multivariable se expresa como sigue:

$$P(A|B_1, B_2, \dots, B_n) = \frac{P(B_1, B_2, \dots, B_n|A) \cdot P(A)}{P(B_1, B_2, \dots, B_n)} \qquad (2)$$

- $P(A|B_1, B_2, \dots, B_n)$ es la probabilidad condicional de que la hipótesis A sea verdadera dado que han ocurrido los eventos $B_1, B_2, \dots, B_n$
- $P(B_1, B_2, \dots, B_n|A)$ es la probabilidad condicional conjunta de que los eventos $B_1, B_2, \dots, B_n$ ocurran dado que la hipótesis A es verdadera.
- $P(A)$ es la probabilidad a priori de la hipótesis A.
- $P(B_1, B_2, \dots, B_n)$ es la probabilidad a priori conjunta de los eventos $B_1, B_2, \dots, B_n$

Este teorema es especialmente útil en situaciones donde hay dependencias entre múltiples variables aleatorias, y se quiere actualizar la probabilidad de una hipótesis dado un conjunto de observaciones.

C. Algoritmo de Bayes

Los algoritmos de Naive Bayes son una clase de algoritmos de clasificación de aprendizaje automático o machine learning. Están basados en el teorema de bayes.
Estos algoritmos de clasificación asumen que el efecto de una característica en particular en una clase es independiente de otras características.
Si estas características son interdependientes, estas características se consideran de forma independiente. Esta suposición se denomina independencia condicional de clase y simplifica en gran medida las operaciones que realizan los códigos de computadora, por lo que se considera una técnica ingenua.
Retomando la ecuación 1:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \qquad (3)$$

- $P(B) =$ Es la probabilidad previa de que la hipótesis B sea cierta. Es la probabilidad previa de B.
- $P(A) =$ Probabilidad previa los datos previos. Está basada en los datos de eventos pasados.
- $P(B|A) =$ Es la probabilidad de la hipótesis B dados los datos A.
- $P(A|B) =$ Es la probabilidad de los datos A dado que la hipótesis B sea cierta.

En caso de que exista una sola característica, el clasificador de Bayes calcula la probabilidad de un evento con los siguientes pasos.

1. Calcular la probabilidad previa para las etiquetas de clase dadas.
2. Determinar la probabilidad con cada atributo ara cada clase
3. Poner estos valores en el teorema de Bayes y calcular la probabilidad posterior.
4. Verificar que clase tiene una probabilidad más alta, dado que la variable de entrada pertenece a la clase de probabilidad más alta.

D. Dataset PlayTennis.csv

El dataset PlayTennis.csv es un conjunto de datos de aprendizaje automático que se utiliza para predecir si una persona jugará al tenis o no. El conjunto de datos contiene 14 atributos, cada uno de los cuales representa una condición meteorológica o una variable personal. Los atributos son los siguientes:

- Outlook: Soleado, nublado, lluvioso
- Temperature: Cálido, fresco, frío
- Humidity: Alto, medio, bajo
- Windy: Sí, no
- Play: Sí, no

Outlook	Temperature	Humidity	Windy	PlayTennis
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Tabla 1. Dataset PlayTennis

E. Dataset Tienda.csv

Para seguir comprobando el funcionamiento del algoritmo de Bayes, se ha propuesto la resolución del siguiente problema.

Una tienda que vende aparatos electrónicos y accesorios de vestimenta registra en su base de datos los productos que vende. Los datos que guarda de sus productos son la categoría a la que pertenecen, el producto del que se trata, su existencia en stock, si se encuentra actualmente con una oferta de descuento y si el producto es popular o no en sus ventas. Los datos los podemos visualizar con mayor detalle en la tabla 2.

Categoría	***Producto***	***En Stock***	***Oferta***	***Popularidad***
Electrónicos	Laptop	True	No	No Popular
Electrónicos	Laptop	False	Yes	Popular
Electrónicos	Laptop	True	Yes	Popular
Electrónicos	Smartphone	True	No	Popular
Electrónicos	Smartphone	False	Yes	Popular
Electrónicos	Smartphone	False	No	No Popular
Electrónicos	Tablet	False	Yes	Popular
Electrónicos	Auriculares	True	Yes	Popular
Accesorios	Mochila	True	Yes	Popular
Accesorios	Mochila	True	No	No Popular
Accesorios	Gafas	False	Yes	Popular
Accesorios	Gorra	True	No	No Popular
Accesorios	Gorra	False	Yes	No Popular
Accesorios	Gafas	False	No	No Popular

Tabla 2. Dataset Tienda.

Estos datos han sido recopilados por la tienda en base a las ultimas ventas que ha realizado. Algunos de sus productos ya no se encuentran en stock aunque son muy populares, por el contrario existen productos que aún están en stock pero no son tan vendidos. Para tomar mejores decisiones en cuanto a la adquisición de más productos en stock y ofrecer ofertas a los clientes, la tienda desea conocer cuáles son los productos que tienen más probabilidad de ser comprados según su popularidad.

II. Desarrollo

A. Dataset PlayTennis.csv

Aplicando el algoritmo descrito con anterioridad, obtenemos:

Temp	Play
Mild	yes
Mild	yes
Mild	yes
Mild	yes
Mild	no
Mild	no
Hot	yes
Hot	yes
Hot	no
Hot	no
Cool	yes
Cool	yes
Cool	yes
Cool	no

→

Frecuencia

Temp	No	Yes
Mild	2	4
Hot	2	2
Cool	1	3
Total	5	9

↓

Verosimilitud 1

Temp	No	Yes		
Mild	2	4	=6/14	0.42
Hot	2	2	=4/14	0.28
Cool	1	3	=4/14	0.28
Total	5	9		
	=5/14	=9/14		
	0.35	0.64		

↓

Verosimilitud 2

Temp	No	Yes	Posterior Probabilidad No	Posterior Probabilidad yes
Mild	2	4	2/5=0.4	4/9=0.44
Hot	2	2	2/5=0.4	2/9=0.22
Cool	1	3	1/5=0.2	3/9=0.33
Total	5	9		

a) Ejemplo 1:

Ahora queremos calcular la probabilidad de jugar cuando la temperatura está "Mild"

$$P(Yes|Mild) = \frac{P(Mild|Yes) \cdot P(Yes)}{P(Mild)}$$

Primero, calculamos las probabilidades apriori:

$$P(Mild) = \frac{6}{14} = 0.42$$

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

Poner las probabilidades a priori y a posteriori en la ecuación

$$P(Yes|Mild) = \frac{0.44 \cdot 0.64}{0.42} = \mathbf{0.67}$$

Ahora comparándolo cuando no se juega cuando la temperatura está "Mild":

$$P(No|Mild) = \frac{P(Mild|No) \cdot P(No)}{P(Mild)}$$

Primero, calculamos las probabilidades apriori:

$$P(Mild) = \frac{6}{14} = 0.42$$

$$P(No) = \frac{9}{14} = 0.35$$

Poner las probabilidades a priori y a posteriori en la ecuación

$$P(No|Mild) = \frac{0.4 \cdot 0.35}{0.42} = \mathbf{0.33}$$

b) Ejemplo 2

Realizando las mismas operaciones, pero con temperatura en "Hot" obtenemos:

$$P(Yes|Hot) = \frac{P(Hot|Yes) \cdot P(Yes)}{P(Hot)}$$

Primero, calculamos las probabilidades apriori:

$$P(Hot) = \frac{4}{14} = 0.28$$

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

Poner las probabilidades a priori y a posteriori en la ecuación

$$P(Yes|Hot) = \frac{0.22 \cdot 0.64}{0.28} = \mathbf{0.502}$$

Ahora con no jugar:

$$P(No|Hot) = \frac{0.4 \cdot 0.35}{0.28} = \mathbf{0.5}$$

B. Dataset Tienda.csv

Categoría	Popularidad
Electronicos	No Popular
Electronicos	Popular
Electronicos	Popular
Electronicos	Popular
Electronicos	Popular
Electronicos	Popular
Electronicos	Popular
Electronicos	No Popular
Accesorios	Popular
Accesorios	No Popular
Accesorios	No Popular
Accesorios	No Popular
Accesorios	No Popular
Accesorios	Popular

→

Frecuencia

Categoría	No P	P
Electronicos	2	6
Accesorios	4	2

↓

Verosimilitud 1

Categoría	No P	P		
Electronicos	2	6	=8/14	0.57
Accesorios	4	2	=6/14	0.42
Total	6	8		
	=6/14	=8/14		
	0.42	0.57		

Verosimilitud 2

Categoria	No P	P	Posterior Probabilidad No P	Posterior Probabilidad P
Electronicos	2	6	2/6=0.3	6/8=0.75
Accesorios	4	2	4/6=0.6	2/8=0.25
Total	6	8		

a) Ejemplo 1:
Calculemos la probabilidad de que tan popular son los electrónicos

$$P(Po|Electro) = \frac{P(Electro|Po) \cdot P(Po)}{P(Electro)}$$

$$P(Electro) = \frac{8}{14} = 0.57$$

$$P(Po) = \frac{8}{14} = 0.57$$

$$P(Po|Electro) = \frac{0.75 \cdot 0.57}{0.57} = \mathbf{0.75}$$

Ahora el caso de no popular

$$P(No\ P|Electro) = \frac{P(Electro|No\ P) \cdot P(No\ P)}{P(Electro)}$$

$$P(Electro) = \frac{8}{14} = 0.57$$

$$P(No\ P) = \frac{6}{14} = 0.42$$

$$P(Po|Electro) = \frac{0.3 \cdot 0.42}{0.57} = \mathbf{0.22}$$

b) Ejemplo 2:
Ahora calculemos la probabilidad de que tan popular son los accesorios

$$P(Po|Acc) = \frac{P(Acc|Po) \cdot P(Po)}{P(Acc)}$$

$$P(Acc) = \frac{6}{14} = 0.42$$

$$P(Po) = \frac{8}{14} = 0.57$$

$$P(Po|Acc) = \frac{0.25 \cdot 0.57}{0.42} = \mathbf{0.33}$$

Ahora el caso de no popular

$$P(No\ P|Acc) = \frac{P(Acc|No\ P) \cdot P(No\ P)}{P(Acc)}$$

$$P(Acc) = \frac{6}{14} = 0.42$$

$$P(Acc) = \frac{6}{14} = 0.42$$

$$P(No\ P|Acc) = \frac{0.6 \cdot 0.42}{0.42} = \mathbf{0.6}$$

III. RESULTADOS

A. DATASET PLAYTENNIS.CSV

Aplicando los datos al algoritmo de bayes obtenemos como resultado:

Hipótesis → Temp: Mild

$$yes \rightarrow\ 0.66$$

$$No \rightarrow 0.33$$

Lo que es muy parecido cuando realizamos las operaciones a mano. Los resultados nos indican hay una mayor probabilidad de que la variable objetivo sea "yes" en comparación con "no", lo que nos dice que se puede jugar mayoritariamente

Sin embargo, cuando aplicamos las mismas operaciones a:

Hipótesis → Temp: Hot

$$yes \rightarrow\ 0.54$$

$$No \rightarrow 0.45$$

A diferencia de los resultados anteriores podemos ver que las probabilidades son muy similares, es muy mínima la diferencia entre jugar y no jugar.

Aplicando el algoritmo con múltiples variables, podemos agregar a

Hipótesis → Temp: Mild y Humidity: Normal

$$yes \rightarrow\ 0.72$$

$$No \rightarrow 0.27$$

Observamos como la probabilidad de jugar es mayor a diferencia de solamente aplicar un evento simple, lo podemos comparar aplicando:

Hipótesis → Temp: Mild y Humidity: High

$$yes \rightarrow\ 0.37$$

$$No \rightarrow 0.62$$

Donde vemos un claro cambio, donde con humedad alta no es recomendable jugar.

Podemos aplicar una tercera condición:

Hipótesis → Temp: Mild , Humidity: Normal y Windy: falso

$$yes \rightarrow \ 0.79$$

$$No \rightarrow 0.20$$

Donde observamos que mantenemos la lógica que cuando la temperatura esta templada, la humedad es normal y no hay viento entonces es viable que se juegue y podemos comparar el caso contrario:

Hipótesis → Temp: Mild , Humidity: Normal y Windy: true

$$yes \rightarrow \ 0.62$$

$$No \rightarrow 0.37$$

Donde observamos una reducción en la probabilidad de si, sin embargo, sigue siendo mayor a diferencia de no, también podemos comprobar un caso donde sea mayor la probabilidad de no jugar, cambiando la humedad a alto:

Hipótesis → Temp: Mild , Humidity: Highy Windy: true

$$yes \rightarrow \ 0.27$$

$$No \rightarrow 0.72$$

Donde observamos una posibilidad muy alta en el caso de no jugar

B. Dataset PlayTennis.csv

En el caso de la tienda, probamos el algoritmo con la siguiente hipótesis:

Hipótesis → Categoría: electrónica y Producto: Laptop

$$Popular \rightarrow \ 0.76$$

$$No\ Popular \rightarrow 0.23$$

En este caso la posibilidad de que la compra de un electrodoméstico popular es mayor, por lo que la tienda podría ofrecer más atención a los productos de esta categoría que no son tan populares. Resultados similares a los mostrados en los ejemplos en la sección del dataset de la tienda en el Desarrollo.

Enseguida probamos el algoritmo, pero para una hipótesis que solo involucre a los accesorios.

Hipótesis → Categoría: Accesorios

$$Popular \rightarrow \ 0.39$$

$$No\ Popular \rightarrow 0.60$$

Los resultados son algo diferentes con los esperados, sin embargo, muestra una clara aproximación a los valores calculados.

Podemos probar ahora con otra hipótesis en la que queremos conocer la popularidad de la venta de una mochila:

Hipótesis → Producto: Mochila

$$Popular \rightarrow 0.46$$

$$No\ Popular \rightarrow 0.53$$

En esta ocasión solo tenemos dos ítems de mochila en la que solo hay una que es popular. Incluso, en este ejemplo se puede visualizar a simple vista que la probabilidad de que sea popular y de que no, es del 50%, es decir 0.5. El resultado que sale del código, aunque es diferente, los valores son cercanos.

Jugando con más atributos del datasaet en el código podemos formular la siguiente hipótesis:

Hipótesis → Producto: Laptop y Oferta: yes

$$yes \rightarrow \ 0.84$$

$$No \rightarrow 0.15$$

El algoritmo nos da un valor pequeño para cuando una laptop no es tan popular, lo cual es eserado ues en el dataset solo existe un producto que es Laptop ppero no cuanta con ofertas de descuentos.

Si ahora tomamos en cuenta la cantidad de stock de los roductos de la categoria accseorios, podemos formular la siguiente hipótesis.

Hipótesis → Categoria: Accesorios y EnStock: True

$$yes \rightarrow \ 0.33$$

$$No \rightarrow 0.66$$

IV. Conclusiones

Diego: Con los resultados obtenidos a aplicar el algoritmo de Naive Bayes, se observa cómo la aplicación de probabilidades condicionales y a priori puede proporcionar insights valiosos sobre la relación entre variables predictoras y la variable objetivo. Además, una de las ventajas más importantes de aplicar este algoritmo es que nos permite hacer este tipo de análisis con un buen costo computacional, ya que este tipo de algoritmo tiene una capacidad para realizar predicciones con rapidez lo convierte en una opción atractiva para conjuntos de datos grandes.

Otro punto importante del algoritmo fue su facilidad para interpretar los datos, es decir, es muy claro los resultados obtenidos con las diferentes pruebas, por lo que es fácil ver su utilidad en diferentes áreas como la categorización de correos electrónicos como spam o no spam, diagnóstico médico, y en la clasificación de documentos en temas específicos.

Edkir: A diferencia de otros algoritmos clasificadores, este algoritmo requiere mucha menos poder de cómputo para hacer cálculos. La desventaja es que para que este algoritmo pueda ser escalable a otros casos, va a requerir de algunos cambios para que se pueda ejecutar en diferentes situaciones. Además de que la precisión de los cálculos aun se puede mejorar.

V. BIBLIOGRAFIA

[1] Ligdi G.(15/11/2023) Naive Bayes – Teoría https://aprendeia.com/algoritmo-naive-bayes-machine-learning/

[2] Clasificador bayesiano ingenuo. (2023, 27 de septiembre). Wikipedia, La enciclopedia libre. https://es.wikipedia.org/w/index.php?title=Clasificador_bayesiano_ingenuo&oldid=154106305.

[3] Turing (s. f.) An Introduction to Naive Bayes Algorithm for Beginners https://www.turing.com/kb/an-introduction-to-naive-bayes-algorithm-for-beginners#conditional-probability

[4] Datacamp (2023) Naive Bayes Classification Tutorial using Scikit-learn https://www.datacamp.com/tutorial/naive-bayes-scikit-learn

[5] Surabhi S (14/07/20130) Get Started With Naive Bayes Algorithm: Theory & Implementation https://www.analyticsvidhya.com/blog/2021/01/a-guide-to-the-naive-bayes-algorithm/

[6] Shaier S. (19/02/2019) ML Algorithms: One SD (σ)-Bayesian Algorithms https://towardsdatascience.com/ml-algorithms-one-sd-σ-bayesian-algorithms-b59785da792a

$$p_i = \frac{1}{1 + e^{-(B_0 + B_1 x_{1,i} + \cdots + B_k x_{1,i})}}$$