SISTEMAS INTELIGENTES

Lecture 05:Clasificación con Naïve Bayes

Dr. Edwin Valencia Castillo
Departamento de Sistemas
Facultad de Ingeniería
Universidad Nacional de Cajamarca
2024

Temas a desarrollar

Introducción

Fundamentos teóricos

Tipos de modelos Naïve Bayes

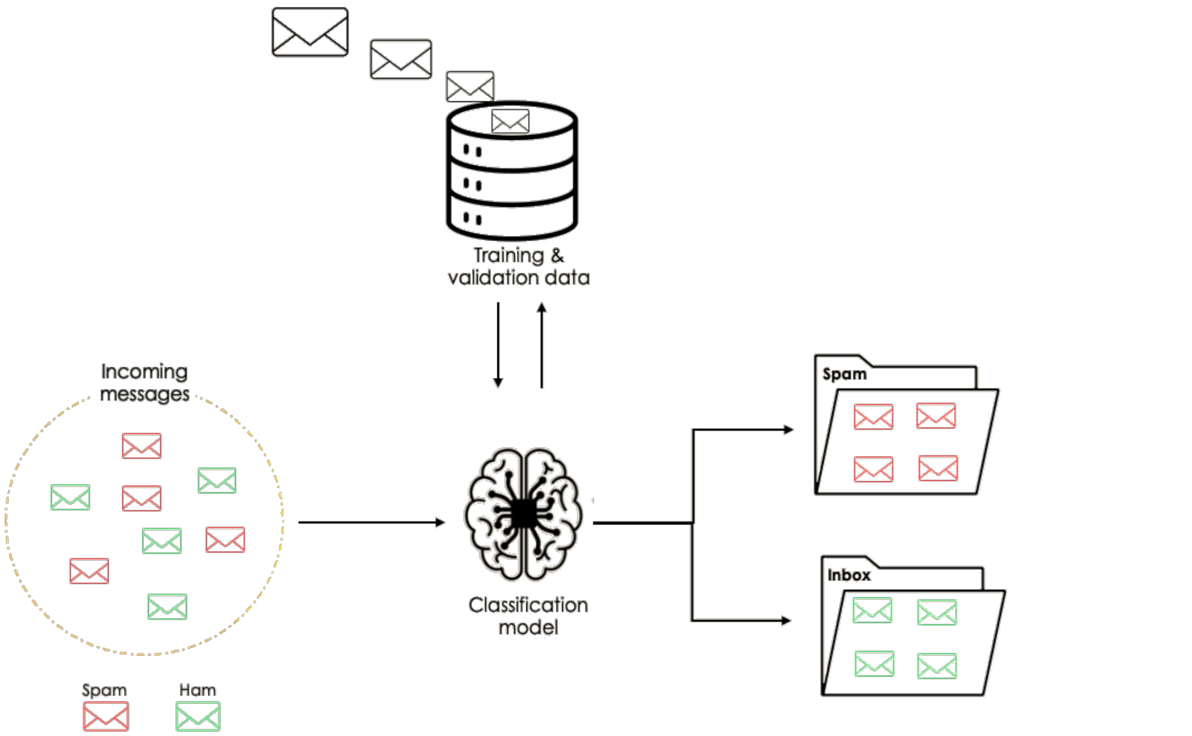
Ejemplo práctico

Introducción

✓ **Recordemos: La clasificación es un método de machine learning supervisado en el que el modelo intenta prever la etiqueta correcta de unos datos de entrada dados.**

➢ **En la clasificación, el modelo se entrena completamente con los datos de entrenamiento y, luego, se evalúa con los datos de prueba antes de utilizarse para realizar previsiones con nuevos datos no vistos.**

Por ejemplo, un algoritmo puede aprender a prever si un determinado correo electrónico es spam o no, como se muestra a continuación.



Introducción

✓ **Objetivos:**

- ➢ Aprender el funcionamiento del clasificador Naïve Bayes.
- ➢ Conocer los diferentes tipos de Naïve Bayes y sus aplicaciones.
- ➢ Implementar un clasificador Naïve Bayes utilizando Python.

✓ **Resultados esperados:**

- ➢ Los estudiantes serán capaces de explicar el teorema de Bayes y la independencia condicional.
- ➢ Podrán aplicar Naïve Bayes a problemas de clasificación utilizando ejemplos prácticos.
- ➢ Evaluarán la eficacia de Naïve Bayes.

Fundamentos Teóricos

- ✓ En este tema se va a introducir el clasificador Naïve Bayes, que está basado en el teorema de Bayes para calcular las probabilidades a posteriori de los eventos a predecir.
- ✓ Probabilidad condicional: Es una medida que cuantifica la probabilidad de que ocurra un evento A dado que ya ha ocurrido otro evento B. Se denota como $P(A|B)$.
- Fórmula: $P(A|B)=P(A\cap B)/P(B)$
- Ejemplo práctico: Probabilidad de que un correo electrónico sea spam dado que contiene la palabra “viagra”.

Fundamentos Teóricos – Teorema de Bayes

- ✓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 solo A.
- ✓El teorema de Bayes es bastante relevante porque relaciona la probabilidad de dos eventos A y B utilizando la dependencia condicional de uno de ellos. Es decir, relaciona la probabilidad de que ocurra el evento A y sabemos de antemano que ha ocurrido B, utilizando la probabilidad de que ocurra el evento B sabiendo que ha ocurrido A.
- ✓Permite relacionar, entre otras cosas, síntomas y enfermedades. Por ejemplo, sabiendo la probabilidad de tener dolor de garganta dado que se conoce que se tiene gripe, se puede obtener la probabilidad de tener gripe dado que se tiene dolor de garganta.

Fundamentos Teóricos – Teorema de Bayes

- ✓ El teorema de Bayes relaciona la comprensión de la probabilidad de aspectos causa-efecto dados los eventos dependientes observados. Un evento dependiente es aquel cuyo resultado se ve afectado por el resultado de otro evento o serie de eventos.
- ✓ Los eventos dependientes son la base del modelado predictivo puesto que se busca obtener la probabilidad de que ocurra un suceso teniendo en cuenta la existencia de una serie de eventos dependientes.
- ✓ Formula: $P(A|B)= P(B|A)\cdot P(A)/ P(B)$

Fundamentos Teóricos – Teorema de Bayes

- ✓Este teorema se puede generalizar para más de dos eventos de la siguiente forma: en primer lugar, se entiende por evento mutuamente excluyente cuando dos resultados diferentes de un evento no pueden ocurrir al mismo tiempo. En el caso además de que los eventos sean exhaustivos, por lo menos uno de ellos tiene que ocurrir.
- ✓De forma matemática, se puede definir el teorema de Bayes para n eventos: Sea $\{A_1, A_2, \ldots, A_i, \ldots, A_n\}$ un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de cero. Sea B un suceso cualquiera del que se conocen las probabilidades condicionales $P(B|A_i)$. Entonces, la probabilidad $P(A_i|B)$ viene dada por la expresión:

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

$P(A_i)$ 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.*
$P(B)$ es la verosimilitud marginal.

Fundamentos Teóricos – Teorema de Bayes

- ✓ **Ejemplo: Supongamos que deseamos estimar la probabilidad de que un mensaje de correo electrónico sea spam. Sin tener evidencias adicionales y, únicamente, por los mensajes previos obtenidos la probabilidad a priori de que un mensaje sea spam es 0,2.**
- ✓ **Ahora bien, si tenemos evidencia de que el mensaje contiene la palabra Viagra, por medio de conocer la probabilidad de que la palabra Viagra haya sido utilizada en mensajes de spam previos (lo cual se conoce como verosimilitud o likelihood) y por medio de la probabilidad de que la palabra Viagra aparezca en cualquier mensaje, lo que se conoce como verosimilitud marginal.**
- ✓ **Aplicando el teorema de Bayes se puede calcular la probabilidad a posteriori y si esta es mayor que 0,5 es más probable que el mensaje sea spam. En la siguiente formula se muestra el cálculo anteriormente descrito.**

$$P(Spam_i|Viagra) = \frac{P(Viagra|Spam) * P(Spam)}{P(Viagra)}$$

El **teorema de Bayes** relaciona la probabilidad de dos o más eventos utilizando la dependencia condicional de cada uno de ellos. Los eventos deben ser dependientes y mutuamente excluyentes

Fundamentos Teóricos – Tablas de probabilidad condicionadas

- ✓ Para obtener cada uno de los componentes de las fórmulas anteriores es necesario construir una tabla de frecuencias que indica el número de veces que el evento aparece en cada una de las situaciones.
- ✓ En nuestro ejemplo de spam, es necesario calcular el número de veces que la palabra Viagra ha aparecido en los mensajes de spam. Esta tabla de frecuencias se utiliza posteriormente para calcular las tablas de verosimilitud o de probabilidad condicionada.
- ✓ Siguiendo con el ejemplo de los mensajes de spam, en el caso hipotético de que tuviéramos la siguiente distribución histórica de 100 mensajes para la palabra Viagra…

Frecuencia	Si	No	Total
Spam	4	16	20
Ham	1	79	80
Total	5	95	100

…Obtendríamos la correspondiente tabla de verosimilitud

Verosimilitud	Si	No	Total
Spam	4/20	16/20	20
Ham	1/80	79/80	80
Total	5/100	95/100	100

Fundamentos Teóricos – Tablas de probabilidad condicionadas

- ✓Con estos datos, para calcular la probabilidad a posteriori de que un mensaje sea spam dado que nos ha llegado la palabra Viagra, tendríamos que hacer el siguiente calculo:

$$P(Spam|Viagra) = [(4/20) * (20/100)] / (5/100) = 0.8$$

- ✓Es decir, con los datos anteriores, la probabilidad de que un correo electrónico que contenga la palabra Viagra sea spam es del 0,8.

Fundamentos Teóricos – Tablas de probabilidad condicionadas

- ✓ Ahora supongamos que deseamos añadir a este cálculo otros términos más comunes que aparecen en los mensajes spam, como pueden ser: money, groceries y unsubscribe.
- ✓ En este caso, tendríamos la siguiente tabla de verosimilitud:

	Viagra (W1)		Money (W2)		Groceries (W3)		Unsubscribe (W4)		
Verosimilitud	Si	No	Si	No	Si	No	Si	No	Total
Spam	4/20	16/20	10/20	10/20	0/20	20/20	12/20	8/20	20
Ham	1/80	79/80	14/80	66/80	8/80	71/80	23/80	57/80	80
Total	5/10	95/10	24/10	76/10	8/100	91/100	35/100	65/100	100

- ✓ De esta forma, si llega un nuevo mensaje que contiene las palabras Viagra y unsubscribe, pero no money ni groceries; utilizando el teorema de Bayes habría que calcular la siguiente formula:

$$P(Spam|Viagra) = \frac{P(Viagra|Spam) * P(Spam)}{P(Viagra)}$$

El cálculo de esta fórmula es computacionalmente costoso, puesto que a medida que se añaden nuevos términos son necesarias grandes cantidades de memoria para almacenar todas las combinaciones

Fundamentos Teóricos – Independencia condicional en el Clasificador Naïve Bayes

- ✓Debido a que el cálculo riguroso de la fórmula del teorema de Bayes, como en el ejemplo anterior, es computacionalmente costoso, el clasificador Naïve Bayes se basa en una modificación sencilla.
- ✓Básicamente asume independencia condicional entre los eventos, a pesar de que si todos los eventos fueran independientes sería imposible predecir ningún evento con los datos observados por otro.
- ✓Formalmente, dos eventos son independientes si el resultado del segundo evento no es afectado por el resultado del primer evento. Si A y B son eventos independientes, la probabilidad de que ambos eventos ocurran es el producto de las probabilidades de los eventos individuales.
- ✓Por otro lado, los eventos dependientes son la base del modelado predictivo, puesto que permiten predecir la presencia de un evento en función de otro. Por ejemplo, la presencia de nubes suele ser un evento predictivo de un día lluvioso, o la presencia de la palaba viagra en un correo electrónico suele ser un evento predictivo de spam.

Fundamentos Teóricos – Independencia condicional en el Clasificador Naïve Bayes

- ✓No obstante, al no poder asumir dependencia condicional por el alto coste computacional, el clasificador Naïve Bayes asume independencia condicional entre los eventos condicionados al mismo valor de la clase. Este hecho es el que le ha dado el adjetivo de Naïve al clasificador.
- ✓En nuestro ejemplo anterior, asumiendo independencia condicional de las palabras para obtener la probabilidad de spam, tendríamos la siguiente formula:

$$P(Spam|W_1 \cap \neg W_2 \cap \neg W_3 \cap W_4) = \frac{P(W_1|Spam)P(\neg W_2|Spam)P(\neg W_3|Spam)P(W_4|Spam) * P(Spam)}{P(W_1)P(\neg W_2)P(\neg W_3)P(W_4)}$$

$$= (4/20) * (10/20) * (20/20) * (12/20) * (20/100) = 0.012$$

Fundamentos Teóricos – Independencia condicional en el Clasificador Naïve Bayes

- ✓Por otro lado, para obtener la probabilidad de ham, tendríamos:

$$P(Ham|W_1 \cap \neg W_2 \cap \neg W_3 \cap W_4) = \frac{P(W_1|Ham)P(\neg W_2|Ham)P(\neg W_3|Ham)P(W_4|Ham) * P(Ham)}{P(W_1)P(\neg W_2)P(\neg W_3)P(W_4))}$$

$$= (1/80) * (66/80) * (71/80) * (23/80) * (80/100) = 0.002$$

- ✓Como 0,012/0,002 = 6, se puede afirmar que es 6 veces más probable que el mensaje sea spam que ham.
- ✓Si queremos calcular la probabilidad de que el mensaje sea spam, sería igual a la verosimilitud de que el mensaje sea spam dividido por la verosimilitud de que sea spam o ham: 0,012/ (0,012 + 0,002) = 0,857
- ✓Análogamente, la probabilidad de ser ham es: 0,002/ (0,002 + 0,012) = 0,143
- ✓Por tanto, podemos estimar que, dadas las palabras del mensaje, hay una probabilidad de 0,857 de que sea spam y de 0,143 de que sea ham y como son eventos mutuamente excluyentes suman 1.

Fundamentos Teóricos – Clasificador Naïve Bayes

- ✓Como hemos comentado previamente, el teorema de Bayes es la base para el clasificador Naïve Bayes. Este clasificador utiliza las probabilidades a priori de los eventos para estimar la probabilidad de eventos futuros por medio del teorema de Bayes.
- ✓Este clasificador utiliza datos históricos o de entrenamiento para calcular la probabilidad observada de cada evento en función de su vector de características.
- ✓Para realizar una predicción, el clasificador es utilizado con datos que tienen la clase desconocida y se utilizan las probabilidades observadas para estimar la clase más probable.

Fundamentos Teóricos – Clasificador Naïve Bayes

- ✓ **La fórmula general del clasificador Naïve Bayes se puede definir de la siguiente manera: la probabilidad del nivel L de la clase C, dada la evidencia proporcionada por las variables de F_1,… F_n, es igual al producto de las probabilidades de cada evidencia condicionada al nivel de la clase, la probabilidad a priori del nivel de la clase y un factor de escala 1/Z que convierte el resultado en probabilidad:**

$$P(C_L|F_1, \dots, F_n) = \frac{1}{Z} p(C_L) \prod_{i=1}^{n} p(F_i|C_L)$$

Este clasificador se utiliza principalmente para **clasificar texto, para detección de intrusos en redes de computadores, diagnósticos médicos,** etc. Por ejemplo, se puede utilizar la frecuencia de las palabras de los correos electrónicos para identificar nuevos correos *spam* en el futuro.

Fundamentos Teóricos – Clasificador Naïve Bayes

- **Combinaciones desconocidas: Supongamos que ahora recibimos un mensaje que contiene las palabras Viagra, groceries, money y unsubscribe, y queremos estimar la probabilidad de que el mensaje sea Viagra. En este caso la verosimilitud de spam es:**

$$(4/20) * (10/20) * (0/20) * (12/20) * (20/100) = 0$$

- **Por otro lado, la verosimilitud de ham es:**

$$(1/80) * (14/80) * (8/80) * (23/80) * (80/100) = 0.00005$$

- **La probabilidad de spam es:**

$$0/(0 + 0.0099) = 0$$

- **Y la probabilidad de ham es:**

$$0.00005 / (0 + 0.00005) = 1$$

- **Este problema sucede cuando un evento nunca ha ocurrido para una o más categorías de las clases. Por ejemplo, si nunca se ha visto el termino groceries en un mensaje spam P(Spam|groceries) = 0.**

Fundamentos Teóricos – Clasificador Naïve Bayes

- ✓La solución es añadir un pequeño número a todas las clases en la tabla, para asegurarse que no existe ninguna combinación con probabilidad de ocurrir igual a 0, esto se conoce con el nombre de estimador de Laplace.
- ✓Por ejemplo, si usamos un valor de 1, la verosimilitud de spam y ham quedaría:

$$(5/24) * (11/24) * (1/24) * (13/24) * (20/100) = 0.0004$$

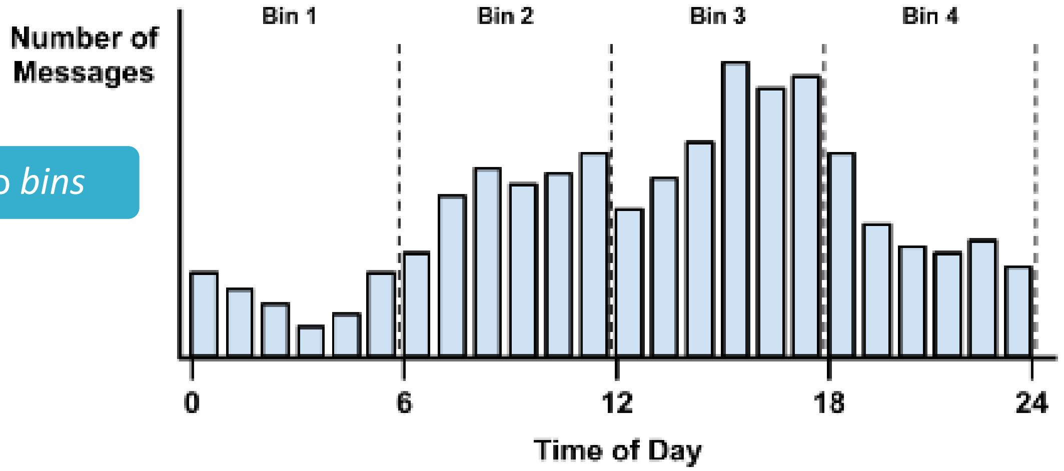
$$(2/84) * (15/84) * (9/84) * (24/84)*(80/100) = 0.0001$$

- ✓Lo que indica que la probabilidad de que el mensaje sea spam es del 0,8 y de que sea ham del 0,2

Fundamentos Teóricos – Clasificador Naïve Bayes con variables numéricas

- ✓ **Debido a que el clasificador Naïve Bayes utiliza tablas de frecuencias para calcular las probabilidades, cada una de las variables utilizada debe de ser categórica y no se pueden utilizar de forma directa variables numéricas.**
- ✓ **Una solución sencilla es discretizar las variables numéricas en N conjuntos, agrupamientos o bins.**
- ✓ **Este método es ideal cuando hay grandes cantidades de datos. Una cuestión importante aquí es considerar el punto de corte óptimo para hacer cada uno de los agrupamientos.**
- ✓ **Una buena solución suele ser explorar los datos para observar los puntos de corte en la distribución de los datos.**
- ✓ **Por ejemplo, el siguiente histograma:**

Sugiere realizar una división en cuatro *bins*



- ✓ **La discretización siempre se traduce en una reducción de la información, ya que la granularidad inicial se reduce. Por tanto, es importante mantener un balance entre el número de bins, puesto que con muy pocas se pierda mucha información y con muchas el proceso es muy costoso.**

Tipos de modelos Naïve Bayes

Multinomial Naïve Bayes: Para datos discretos, como el análisis de texto.

Bernoulli Naïve Bayes: Para datos binarios y presencia/ausencia de caracteristicas.

Gaussian Naïve Bayes: Para datos continuos, asumiendo una distribución normal.

Tipos de modelos Naïve Bayes – Multinomial Naïve Bayes

- ✓ El modelo Multinomial Naïve Bayes es adecuado para datos discretos y es especialmente popular en tareas de clasificación de texto, como la categorización de documentos y el filtrado de spam.
- ✓ En este modelo, se considera la frecuencia de aparición de cada término en el documento.
- ✓ Fórmula
  - ➢ Para una clase C y una característica X_i , la probabilidad condicional se calcula como:

    $P(X_i|C)$=($n_i|C$+α) /(n_C+αN)

    donde:

    $n_i|C$ es el número de veces que la característica X_i aparece en la clase C.

    n_C es el número total de todas las características en la clase C.

    α es un parámetro de suavizado (Laplace smoothing) para evitar probabilidades cero.

    N es el número total de posibles características.
- ✓ Aplicación
  - ➢ Un uso común de Multinomial Naive Bayes es en el filtrado de spam. Cada correo electrónico se representa como un vector de frecuencias de palabras, y el modelo estima la probabilidad de que el correo sea spam o no spam.
- ✓ Ejemplo
  - ➢ Supongamos que estamos clasificando correos electrónicos como spam o no spam basándonos en la frecuencia de las palabras "oferta", "gratis" y "compra". El modelo calcula la probabilidad de cada palabra dada la clase (spam o no spam) y multiplica estas probabilidades junto con la probabilidad previa de cada clase para hacer una predicción.

Tipos de modelos Naïve Bayes – Bernoulli Naïve Bayes

- ✓ El modelo Bernoulli Naïve Bayes es adecuado para datos binarios. En lugar de contar la frecuencia de aparición de características, este modelo se centra en la presencia o ausencia de características.
- ✓ Es útil en problemas donde cada característica se representa como un valor binario (0 o 1).
- ✓ Fórmula
  - ➢ Para una clase C y una característica X_i, la probabilidad condicional se calcula como:

    $P(X_i|C)=(n_i|C+\alpha)/(n_C+2\alpha)$

    donde:

    $n_i|C$ es el número de documentos en la clase C en los que aparece la característica X_i.

    n_C es el número total de documentos en la clase C.

    α es un parámetro de suavizado (Laplace smoothing).
- ✓ Aplicación
  - ➢ Bernoulli Naive Bayes se utiliza en problemas de clasificación de texto donde se considera la presencia o ausencia de palabras. Un ejemplo es la clasificación de documentos en temas específicos (e.g., deportes, política, tecnología) donde cada palabra en el documento se representa como presente (1) o ausente (0).
- ✓ Ejemplo
  - ➢ Si estamos clasificando correos electrónicos como spam o no spam, cada correo se representa como un vector binario donde cada elemento indica la presencia (1) o ausencia (0) de una palabra clave como "oferta" o "gratis". El modelo calcula la probabilidad de cada clase basándose en la presencia o ausencia de estas palabras clave.

Tipos de modelos Naïve Bayes – Gaussian Naïve Bayes

- ✓ El modelo Gaussian Naive Bayes es adecuado para datos continuos. Asume que los valores de las características siguen una distribución normal (gaussiana).
- ✓ Este modelo es útil en problemas donde las características son variables continuas y se pueden aproximar mediante una distribución normal.
- ✓ Fórmula
  - ➢ Para una clase C y una característica X_i, la probabilidad condicional se modela usando la distribución normal:

$$P(X_i|C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} \exp\left(-\frac{(X_i-\mu_C)^2}{2\sigma_C^2}\right)$$

donde:

μ_C es la media de la característica X_i en la clase C.

$\sigma^2{}_C$ es la varianza de la característica X_i en la clase C.

- ✓ Aplicación
  - ➢ Gaussian Naive Bayes se utiliza en problemas donde las características son continuas, como la clasificación de imágenes, análisis de sensores y diagnóstico médico.
- ✓ Ejemplo
  - ➢ Supongamos que estamos clasificando tipos de flores basándonos en características como la longitud y el ancho de los pétalos. Gaussian Naive Bayes estima la probabilidad de cada clase (tipo de flor) asumiendo que estas características siguen una distribución normal en cada clase.

Fortalezas y debilidades

• Simple • Funciona bien con datos ruidosos • Requiere de pocos ejemplos para entrenar • Es fácil obtener la probabilidad estimada para la predicción.	• Se basa en la suposición erronea de variables independientes e igual de importantes • No es lo ideal para datasets con un gran número de variables númericas • Las probabilidades estimadas son menos precisas que las clases predichas

PREGUNTAS??

Dr. Edwin Valencia Castillo
Departamento de Sistemas
Facultad de Ingeniería
Universidad Nacional de Cajamarca
2024

Ejemplo practico: ¿Me conviene comprar casa ó alquilar?

- ✓ En este ejemplo tocaremos el tema sin detalles como como intereses de hipotecas variable/fija, porcentajes, comisiones de bancos, etc, se realizará un planteamiento genérico para obtener resultados y tomar la mejor decisión dada nuestra condición actual.
- ✓ Para este caso se utilizar un algoritmo basado completamente en teoría de probabilidades y obteniendo resultados estadísticos. ¿Será suficiente el Teorema de Bayes para obtener buenas decisiones?
- ✓ Ver jupyter notebook que se alcanza