

Inteligencia Artificial

Juan Pablo Restrepo Uribe

Ing. Biomedico - MSc. Automatización y Control Industrial

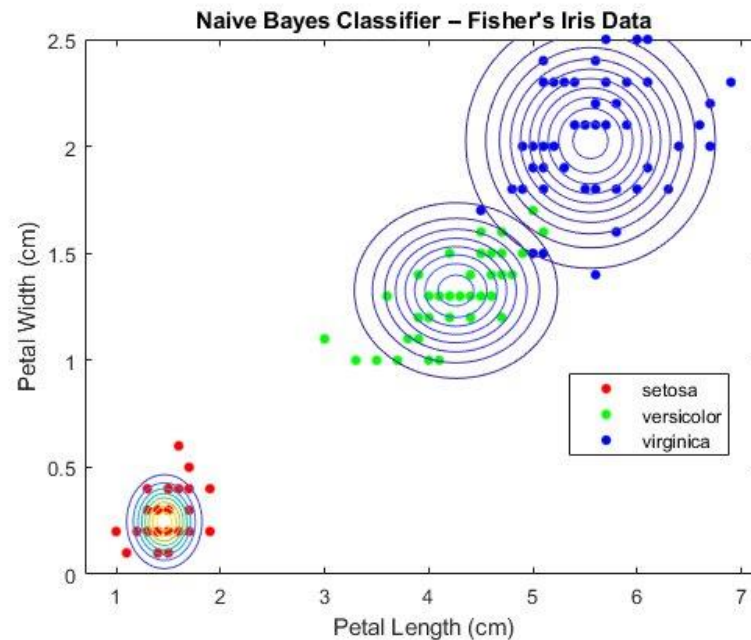
jprestrepo@correo.iue.edu.co

2023

Institución Universitaria de Envigado

Naive Bayes

Se basan en una técnica de clasificación estadística llamada “teorema de Bayes”.



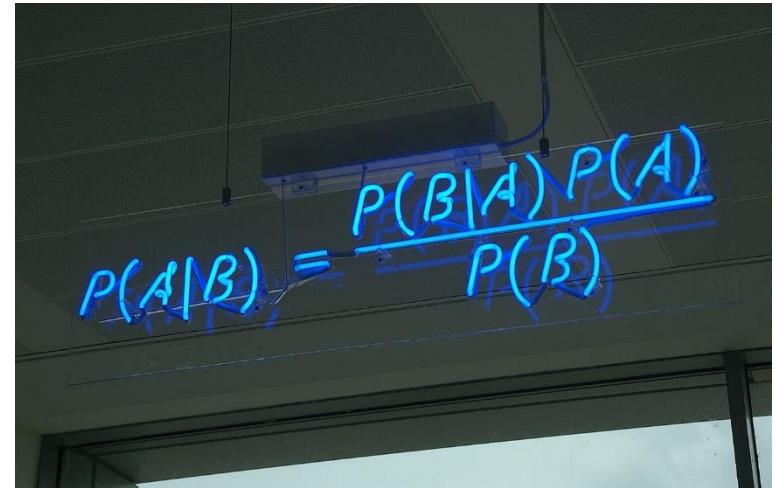
Naive Bayes

En la teoría de la probabilidad, es una proposición planteada por el matemático inglés Thomas Bayes (1702-1761)¹ y publicada póstumamente 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 solo A.

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

Naive Bayes

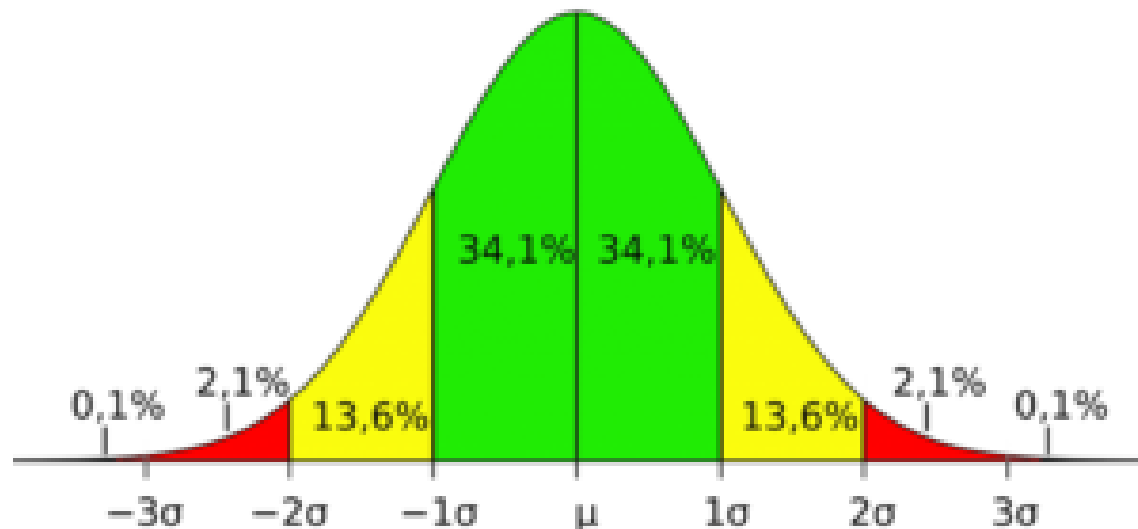
- Estos modelos son llamados algoritmos “Naive”, o “Inocentes” en español. En ellos se asume que las variables predictoras son independientes entre sí.
- Es decir que la presencia de una cierta característica en un conjunto de datos no está en absoluto relacionada con la presencia de cualquier otra característica.



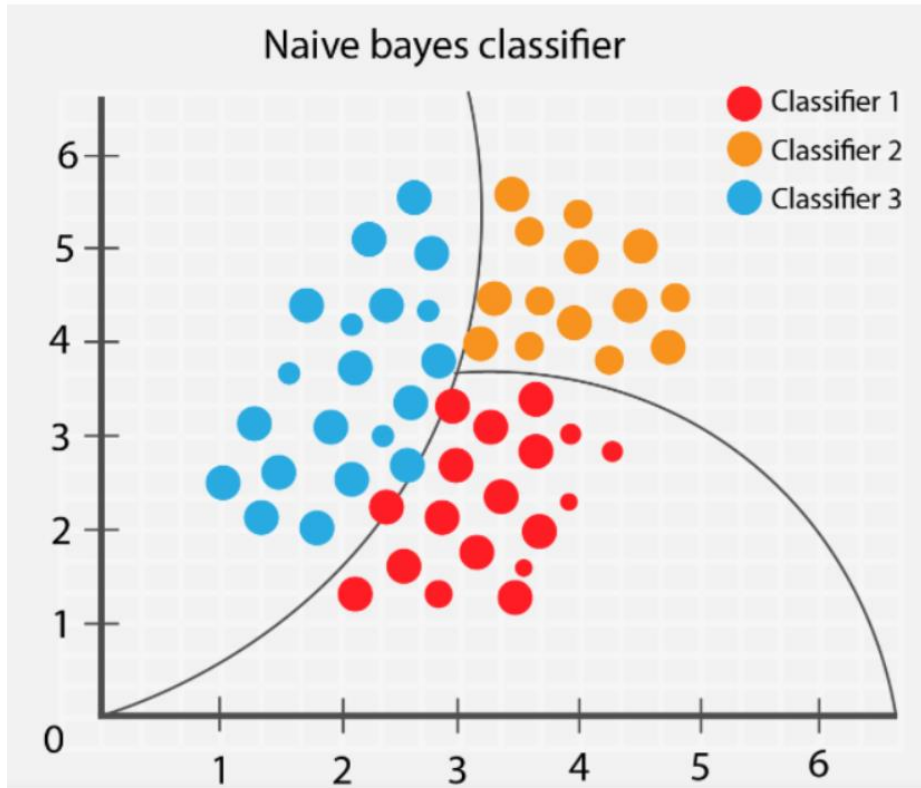
A photograph of a blackboard with the Naive Bayes formula written in blue marker. The formula is
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Probabilidad bayesiana

La probabilidad bayesiana es una de las diferentes interpretaciones del concepto de probabilidad. La interpretación bayesiana de la probabilidad puede ser vista como una extensión de la lógica proposicional que permite razonar con hipótesis, es decir, las proposiciones cuya veracidad o falsedad son inciertas.



Pasos para un algoritmo de Naive Bayes



- Convertir el conjunto de datos en una tabla de frecuencias.
- 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.

Ventajas de Naive Bayes

- Una manera fácil y rápida de predecir clases, para problemas de clasificación binarios y multiclase.
- En los casos en que sea apropiada una presunción de independencia, el algoritmo se comporta mejor que otros modelos de clasificación, incluso con menos datos de entrenamiento.
- El desacoplamiento de las distribuciones de características condicionales de clase significan que cada distribución puede ser estimada independientemente como si tuviera una sola dimensión. Esto ayuda con problemas derivados de la dimensionalidad y mejora el rendimiento.

Desventajas de Naive Bayes

- 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.
- La presunción de independencia Naive muy probablemente no reflejará cómo son los datos en el mundo real.
- Cuando el conjunto de datos de prueba tiene una característica que no ha sido observada en el conjunto de entrenamiento, el modelo le asignará una probabilidad de cero y será inútil realizar predicciones. Uno de los principales métodos para evitar esto, es la técnica de suavizado, siendo la estimación de Laplace una de las más populares.

Proyecto de aplicación

Actualmente, una de las aplicaciones principales de Machine Learning es la detección de spam. Casi todos los servicios de email más importantes proporcionan un detector de spam que clasifica el spam automáticamente y lo envía al buzón de “correo no deseado”.

En este proyecto, desarrollaremos un modelo Naive Bayes que clasifica los mensajes SMS como spam o no spam ('ham' en el proyecto). Se basará en datos de entrenamiento que le proporcionaremos.

Proyecto de aplicación (Aspectos generales)

Normalmente, en los mensajes de spam se cumple lo siguiente:

- Contienen palabras como: 'gratis', 'gana', 'ganador', 'dinero' y 'premio'.
- Tienden a contener palabras escritas con todas las letras mayúsculas y tienden al uso de muchos signos de exclamación.
- Esto es un problema de clasificación binaria supervisada, ya que los mensajes son o 'Spam' o 'No spam' y alimentaremos un conjunto de datos etiquetado para entrenar el modelo.

Proyecto de aplicación (pasos)

- Entender el conjunto de datos
- Procesar los datos
- División del conjunto de datos (Dataset) en los grupos de entrenamiento y pruebas
- Aplicar “Bag of Words” (BoW) para procesar nuestro conjunto de datos
- Implementación de Naive Bayes con Sci-kit Learn
- Evaluación del modelo
- Conclusión

Bag of Words

Es un término usado para especificar los problemas que tiene una colección de datos de texto que necesita ser procesada. La idea es tomar un fragmento de texto y contar la frecuencia de las palabras en el mismo.

	about	bird	heard	is	the	word	you
About the bird, the bird, bird bird bird	1	5	0	0	2	0	0
You heard about the bird	1	1	1	0	1	0	1
The bird is the word	0	1	0	1	2	1	0

Bag of Words

['Hello, how are you!', 'Win money, win from home.', 'Call me now', 'Hello, Call you tomorrow?']

	are	call	from	hello	home	how	me	money	now	tomorrow	win	you
0	1	0	0	1	0	1	0	0	0	0	0	1
1	0	0	1	0	1	0	0	1	0	0	2	0
2	0	1	0	0	0	0	1	0	1	0	0	0
3	0	1	0	1	0	0	0	0	0	1	0	1

Gaussian Naive Bayes¶

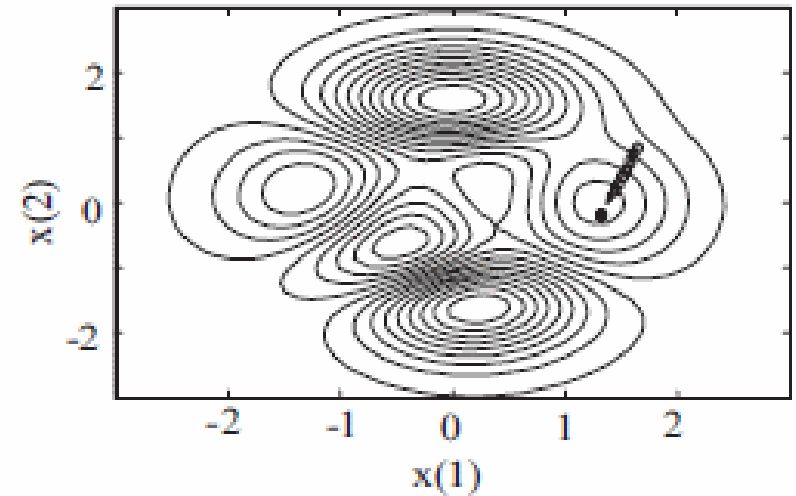
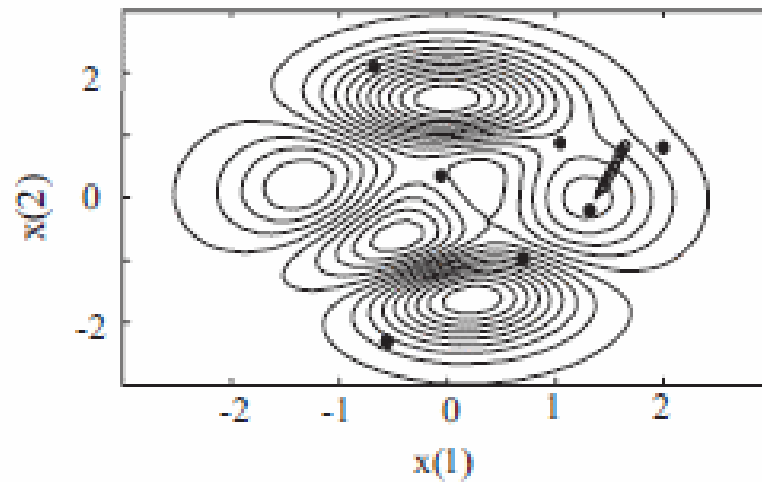
$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp \left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2} \right)$$

Multinomial Naive Bayes

Es una de las dos variantes clásicas de Naive Bayes utilizadas en la clasificación de texto (donde los datos se representan típicamente como conteos de vectores de palabras, aunque también se sabe que los vectores tf-idf funcionan bien en la práctica)

Word	P(word Sports)	P(word Not Sports)
a	$\frac{2 + 1}{11 + 14}$	$\frac{1 + 1}{9 + 14}$
very	$\frac{1 + 1}{11 + 14}$	$\frac{0 + 1}{9 + 14}$
close	$\frac{0 + 1}{11 + 14}$	$\frac{1 + 1}{9 + 14}$
game	$\frac{2 + 1}{11 + 14}$	$\frac{0 + 1}{9 + 14}$

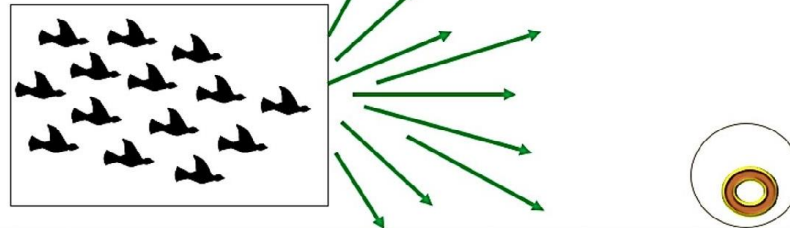
Particle Swarm Optimization (PSO)



Particle Swarm Optimization (PSO)

Pertenece a las técnicas denominadas optimización inteligente y se clasifica como un algoritmo estocástico de optimización basado en población. A esta clasificación igualmente pertenecen los Algoritmos Genéticos (AG).

- All the Birds do not know where food is, but they know how far they are in each iteration.



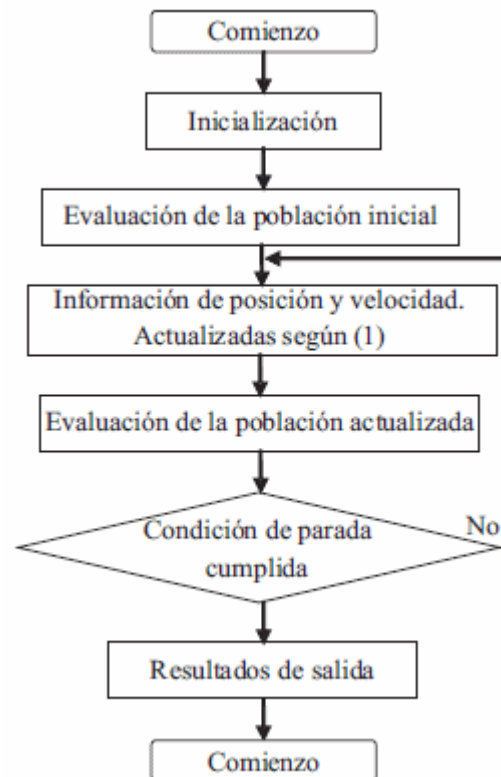
Este algoritmo pretende representar el proceso natural de comunicación grupal para compartir conocimiento individual cuando grupos de animales se desplazan, migran o cazan. Si un miembro detecta un camino deseable para desplazarse, el resto de la colonia lo sigue inmediatamente.

Particle Swarm Optimization (PSO)

El comportamiento animal es imitado por partículas con ciertas posiciones y velocidades en un espacio de búsqueda, donde la población es llamada swarm, y cada miembro del swarm es llamado partícula.

La población inicial se determina aleatoriamente y cada partícula se desplaza a través del espacio de búsqueda y recuerda la mejor posición que ha encontrado. Cada partícula comunica las buenas posiciones a las demás y dinámicamente ajustan su propia posición y su velocidad con base en las buenas posiciones. La velocidad se ajusta con el comportamiento histórico de las partículas

Particle Swarm Optimization (PSO)



Particle Swarm Optimization (PSO)

[1/100] $w:0.800 - c_1:3.500 - c_2:0.500$

