

Algoritmos Naive Bayes

Naïve Bayes o el **Ingenuo** Bayes es uno de los algoritmos **más simples** y poderosos para la clasificación basado en el Teorema de Bayes con **una suposición de independencia entre los predictores**. Naive Bayes es **fácil de construir** y particularmente útil para conjuntos de **datos muy grandes**.

El clasificador Naive Bayes asume que el efecto de una **característica particular en una clase es independiente de otras características**. Por ejemplo, un solicitante de préstamo es deseable o no dependiendo de sus ingresos, historial de préstamos y transacciones anteriores, edad y ubicación. Incluso si estas características son interdependientes, estas características se consideran de forma independiente. Esta suposición simplifica la computación, y por eso se considera ingenua. Esta suposición se denomina independencia condicional de clase.

La formula del teorema de Bayes es la siguiente:

$$P(h | D) = \frac{P(D | h)P(h)}{P(D)}$$

Donde:

$P(h)$: es la probabilidad de que la hipótesis h sea cierta (independientemente de los datos). Esto se conoce como la probabilidad previa de h .

$P(D)$: probabilidad de los datos (independientemente de la hipótesis). Esto se conoce como probabilidad previa.

$P(h|D)$: es la probabilidad de la hipótesis h dada los datos D . Esto se conoce como la probabilidad posterior.

$P(D|h)$: es la probabilidad de los datos d dado que la hipótesis h era cierta. Esto se conoce como probabilidad posterior.

Algoritmos Naive Bayes

En caso de que se tenga una sola característica, el clasificador Naive Bayes calcula la probabilidad de un evento en los siguientes pasos:

Paso 1: calcular la probabilidad previa para las **etiquetas de clase** dadas.

Paso 2: determinar la **probabilidad de probabilidad con cada atributo para cada clase**.

Paso 3: poner estos valores en el teorema de Bayes y **calcular la probabilidad posterior**.

Paso 4: **ver qué clase tiene una probabilidad más alta**, dado que la variable de entrada pertenece a la clase de probabilidad más alta.

Ventajas

Es **fácil y rápido** predecir la clase de conjunto de datos de prueba. También funciona bien en la **predicción multiclase**.

Cuando se mantiene la **suposición de independencia**, un clasificador Naive Bayes **funciona mejor en comparación con otros modelos** como la Regresión Logística y se necesitan **menos datos de entrenamiento**.

Funciona bien en el caso de variables de **entrada categóricas comparada con variables numéricas**.

Desventajas

Si la variable categórica tiene una **categoría en el conjunto de datos de prueba**, que **no se observó en el conjunto de datos de entrenamiento**, el modelo **asignará una probabilidad de 0 y no podrá hacer una predicción**. Esto se conoce a menudo como **frecuencia cero**. Para resolver esto, podemos utilizar la técnica de alisamiento.

Otra limitación de Naive Bayes es la **asunción de predictores independientes**. En la vida real, es casi **imposible que obtengamos un conjunto de predictores que sean completamente independientes**.

Naive Bayes **es el algoritmo más sencillo y potente**. A pesar de los significativos avances de Machine Learning en los últimos años, ha demostrado su valía. Se ha implementado con éxito en muchas aplicaciones, desde el **análisis de texto hasta los motores de recomendación**.