

Machine Learning

Naïve Bayes

Clasificador Bayesiano Ingenuo

- **Naïve Bayes** (Bayes Ingenuo) es un **método de clasificación** usado, principalmente, para tareas de clasificación de documentos (texto).
- Es un **método cualitativo**: representa como **se relacionan los atributos** de forma causal o señalando correlación entre ellos.
- Es un **método cuantitativo**: da **una medida probabilística** de la importancia de los atributos.

Clasificador Bayesiano Ingenuo

Propiedades de este método bayesiano:

- Cada ejemplo observado **modifica la probabilidad** de que la hipótesis formulada sea correcta, aumentándola o disminuyéndola.
- Es **robusto al posible ruido** presente en los ejemplos de entrenamiento (datos incompletos o erróneos).
- Permite que la predicción **considere el conocimiento a priori** o conocimiento del dominio en forma de probabilidades.

Clasificadores bayesianos. El algoritmo Naïve Bayes. Constantino Malagón Luque.

Derivación del Clasificador Bayesiano Ingenuo

Sea $\mathbf{x} = (x_1, x_2, \dots, x_n)$ una instancia del conjunto de datos la cual tiene asignada una de k posibles clases. El clasificador Bayesiano Ingenuo (BI) busca determinar su clase correcta, que se puede representar como:

$$\operatorname{argmax}_k p(C_k | x_1, \dots, x_n)$$

¿Cómo se puede representar la expresión anterior utilizando el teorema de Bayes?

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

Derivación del Clasificador Bayesiano Ingenuo

Sea $\mathbf{x} = (x_1, x_2, \dots, x_n)$ una instancia del conjunto de datos la cual tiene asignada una de k posibles clases. El clasificador Bayesiano Ingenuo (BI) busca determinar su clase correcta, que se puede representar como:

$$\operatorname{argmax}_k p(C_k | x_1, \dots, x_n)$$

Solución:

$$p(C_k | \mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

Derivación del Clasificador Bayesiano Ingenuo

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

El numerador de la expresión anterior se puede expresar como una función de probabilidad conjunta:

$$p(C_k, x_1, \dots, x_n)$$

¿Por qué?

Derivación del Clasificador Bayesiano Ingenuo

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

El numerador de la expresión anterior se puede expresar como una función de probabilidad conjunta:

$$p(C_k, x_1, \dots, x_n)$$

¿Por qué?

$$\begin{aligned} P(X = x, Y = y) &= P(Y = y|X = x)P(X = x) \\ &= P(X = x|Y = y)P(Y = y) \end{aligned}$$

Derivación del Clasificador Bayesiano Ingenuo

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

Expandir recursivamente empezando por x_1

$$p(C_k, x_1, \dots, x_n)$$

Considerando que

$$\begin{aligned} P(X = x, Y = y) &= P(Y = y|X = x)P(X = x) \\ &= P(X = x|Y = y)P(Y = y) \end{aligned}$$

Derivación del Clasificador Bayesiano Ingenuo

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

Expandiendo

$$p(C_k, x_1, \dots, x_n)$$

se tiene que:

$$p(C_k, x_1, \dots, x_n) = p(x_1|x_2, \dots, x_n, C_k) \cdot p(x_2, \dots, x_n, C_k)$$

$$= p(x_1|x_2, \dots, x_n, C_k) \cdot p(x_2|x_3, \dots, x_n, C_k) \cdot p(x_3, \dots, x_n, C_k)$$

⋮

$$= p(x_1|x_2, \dots, x_n, C_k) \cdot \dots \cdot p(x_{n-1}|x_n, C_k) \cdot p(x_n|C_k) \cdot p(C_k)$$

Derivación del Clasificador Bayesiano Ingenuo

$$p(x_1|x_2, \dots, x_n, C_k) \cdot \dots \cdot p(x_{n-1}|x_n, C_k) \cdot p(x_n|C_k) \cdot p(C_k)$$

Si suponemos que cada característica es independiente una de otra, i.e., x_i es independiente de x_j , se tiene que:

$$p(x_i|x_{i+1}, \dots, x_n, C_k) = p(x_i|C_k)$$

Por lo que

$$\begin{aligned} p(C_k|x_1, \dots, x_n) &\propto p(C_k, x_1, \dots, x_n) \\ &\propto p(C_k)p(x_1|C_k) \cdot \dots \cdot p(x_n|C_k) \\ &\propto p(C_k) \prod_{i=1}^n p(x_i|C_k) \end{aligned}$$

¿Dónde se encuentra el paso ingenuo?

Derivación del Clasificador Bayesiano Ingenuo

$$p(C_k|x_1, \dots, x_n) \propto p(C_k) \prod_{i=1}^n p(x_i|C_k)$$

Del desarrollo anterior se obtiene la probabilidad de pertenencia a una determinada clase, por lo que simplemente se opta por elegir la de mayor valor:

$$\operatorname{argmax}_k p(C_k) \prod_{i=1}^n p(x_i|C_k)$$

Derivación del Clasificador Bayesiano Ingenuo

Si recuerdan,

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})}$$

¿Qué pasa con el denominador?

Es constante, por lo que se omite su contribución.

¿Cómo se determinan parámetros

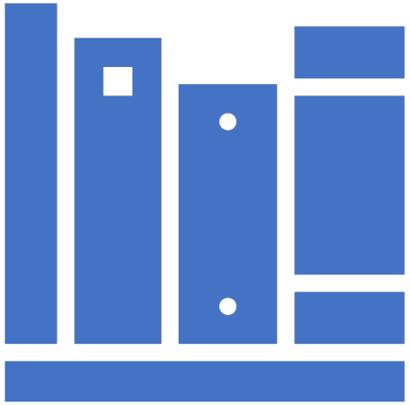
- $p(x_i)$ se determina contando las veces que ocurre la característica x_i en el conjunto de entrenamiento y dividiéndolo por el total de elementos en dicho conjunto.
- $p(x_i | C_k)$ sería el conteo de observaciones de x_i en la clase k en el conjunto de entrenamiento, dividiéndolo entre la cardinalidad de dicho conjunto.

Tarea

Investigar cómo funciona el algoritmo Bayesiano Ingenuo Gaussiano (*Gaussian Naïve Bayes*):

- Diferencias con el método base de Bayes Ingenuo.
- Planteamiento teórico y consideraciones adicionales.
- Cómo se ven esos cambios en la formulación.
- Extras: interpretaciones gráficas.

Ojo: Buscar fuentes bibliográficas adecuadas para desarrollar su investigación. Cualquier intento de sólo traducir + copiar + pegar, en lugar de **desarrollar el tema de forma adecuada**, hará que **la tarea se considere nula**.



Final de la presentación

¡Gracias!