

Naïve Bayes

Jesús D. Hernández Londoño¹

¹Universidad de Puerto Rico. Departamento de Matemáticas.
Recinto Mayagüez.

14 de septiembre de 2020



Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo
- 4 Corrección de laplace
- 5 Ventajas y Desventajas
- 6 Referencias



Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo
- 4 Corrección de laplace
- 5 Ventajas y Desventajas
- 6 Referencias



Naïve Bayes

- Es un Modelo de Clasificación basado en la probabilidad Bayesiana.
- Es un caso particular de Redes Bayesianas.
- Las variables predictoras son independientes.
- Dado un vector de entradas $(X_1 = u_1, X_2 = u_2, \dots, X_m = u_m)$, calcular la probabilidad posterior $P(Y = v_i | X_1, X_2, \dots, X_m)$ sea la mayor posible.



Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo
- 4 Corrección de laplace
- 5 Ventajas y Desventajas
- 6 Referencias



Predicciones

$$\begin{aligned} Y^{Predict} &= \operatorname{argmax} (P(Y = v | X_1, X_2, \dots, X_m)) \\ &= \operatorname{argmax} (P(X_1, X_2, \dots, X_m | Y = v) P(Y = v)) \\ &= \operatorname{argmax} (P(X_1 = u_1 | Y = v) \dots P(X_m = u_m | Y = v) P(Y = v)) \\ &= \operatorname{argmax} \left(P(Y = v) \prod_{j=1}^m P(X_j = u_j | Y = v) \right) \end{aligned}$$

Muchas predictoras en el producto pueden producir underflow, entonces mejor usar logaritmos.

$$Y^{Predict} = \operatorname{argmax} \left(\log P(Y = v) \sum_{j=1}^m \log P(X_j = u_j | Y = v) \right)$$



X Discreta

$$P(X_j = u_j | Y = v) = \frac{\text{\#Records con } X_j = u_j \text{ en la clase } V}{\text{\#Records en la clase } V \text{ en } Y}$$

X Continua

- Discretización.
- Asumir una distribución para cada predictora, en general, Gaussiana con media y varianza estimada de los datos en la clase V .



Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo**
- 4 Corrección de laplace
- 5 Ventajas y Desventajas
- 6 Referencias



Ejemplo 1

A que clase será asignada el registro ($X_1 = 0, X_2 = 0, X_3 = 1$)?

| X1 | X2 | X3 | Y |
|----|----|----|---|
| 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 |
| 1 | 1 | 0 | 0 |
| 0 | 0 | 1 | 1 |
| 1 | 1 | 1 | 1 |
| 0 | 0 | 1 | 1 |
| 1 | 1 | 0 | 1 |



Ejemplo 1

$$\begin{aligned} &P(X_1 = 0, X_2 = 0, X_3 = 1 | Y = 0)P(Y = 0) \\ &= P(X_1 = 0 | Y = 0)P(X_2 = 0 | Y = 0)P(X_3 = 1 | Y = 0)P(Y = 0) \\ &= \left(\frac{2}{3}\right) \left(\frac{1}{3}\right) \left(\frac{1}{3}\right) \left(\frac{3}{7}\right) = \frac{2}{63} \end{aligned}$$

$$\begin{aligned} &P(X_1 = 0, X_2 = 0, X_3 = 1 | Y = 1)P(Y = 1) \\ &= P(X_1 = 0 | Y = 1)P(X_2 = 0 | Y = 1)P(X_3 = 1 | Y = 1)P(Y = 1) \\ &= \left(\frac{2}{4}\right) \left(\frac{2}{4}\right) \left(\frac{3}{4}\right) \left(\frac{4}{7}\right) = \frac{3}{28} \end{aligned}$$

Como $\frac{2}{63} < \frac{3}{28}$ la $(X_1 = 0, X_2 = 0, X_3 = 1)$ será asignado a la clase 1



Ejemplo 2

A que clase será asignada el registro ($X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 4,25$)?

| X1 | X2 | X3 | X4 | Y |
|----|----|----|-------|---|
| 0 | 0 | 1 | 3.15 | 0 |
| 0 | 1 | 0 | 8.17 | 0 |
| 1 | 1 | 0 | 5.72 | 0 |
| 0 | 0 | 1 | 7.16 | 1 |
| 1 | 1 | 1 | 9.32 | 1 |
| 0 | 0 | 1 | 12.81 | 1 |
| 1 | 1 | 0 | 15.48 | 1 |



Ejemplo 2

$$\begin{aligned} &P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 4,25 | Y = 0)P(Y = 0) \\ &= P(X_1 = 0 | Y = 0)P(X_2 = 0 | Y = 0)P(X_3 = 1 | Y = 0) \\ &\quad f(X_4 = 4,25 | Y = 0)P(Y = 0) \\ &= \left(\frac{2}{27}\right) dnorm(4,25; 5,68; 2,5102) \left(\frac{3}{7}\right) = 0,0042896 \end{aligned}$$

$$\begin{aligned} &P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 4,25 | Y = 1)P(Y = 1) \\ &= P(X_1 = 0 | Y = 1)P(X_2 = 0 | Y = 1)P(X_3 = 1 | Y = 1) \\ &\quad f(X_4 = 4,25 | Y = 1)P(Y = 1) \\ &= \left(\frac{3}{16}\right) dnorm(4,25; 11,192; 3,68) \left(\frac{4}{7}\right) = 0,0019687 \end{aligned}$$

La $(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 4,25)$ será asignado a la clase 0.



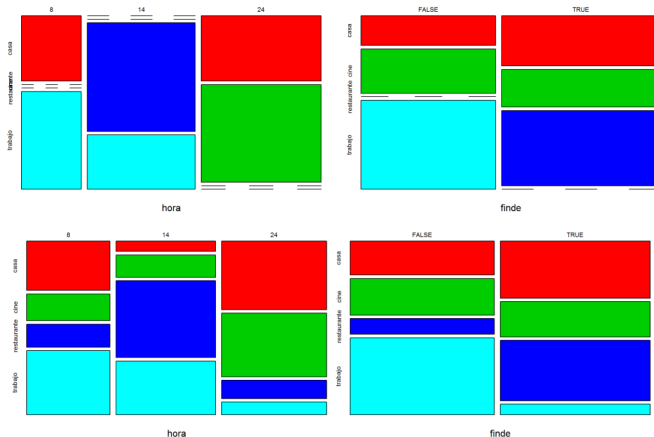
Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo
- 4 Corrección de laplace**
- 5 Ventajas y Desventajas
- 6 Referencias



Corrección de laplace

En muchas ocasiones los datos no contienen muestras de todas las combinaciones de variables posibles.



Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo
- 4 Corrección de laplace
- 5 Ventajas y Desventajas**
- 6 Referencias



- Sencillo.
- Rápido, no tiene problema para trabajar con 10,000 atributos.



- El proceso de discretización parece afectar el rendimiento del clasificador.
- Las probabilidades cero afectan al clasificador, ocurre cuando un valor de una variable en el conjunto de prueba no aparece en el conjunto de entrenamiento. (Corrección de laplace).



Contenido

- 1 Definición
- 2 Predicciones
- 3 Ejemplo
- 4 Corrección de laplace
- 5 Ventajas y Desventajas
- 6 Referencias





Miller J., Forte R.

Mastering Predictive Analytics with R .

Second Edition, Machine learning techniques for advanced models,
2017.



Acuña E.

Data Mining and Machine Learning.

Naïve Bayes



Aprendizaje supervisado en R.

<https://fervilber.github.io/Aprendizaje-supervisado-en-R/>.

Villalba F.

