

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Posgrado en Ciencia e Ingeniería de la Computación

Aprendizaje automatizado

Clasificador Bayesiano Ingenuo

Ayudantes:

Berenice & Ricardo Montalvo Lezama

Profesor:

Gibran Fuentes Pineda

Febrero 2021

- Técnica de clasificación.
- Fundamentado en el Teorema de Bayes.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)}$$

- Aplicaciones: clasificación de textos, análisis de sentimientos y diagnósticos médicos.

- Construcción del clasificador:

- Asumir una distribución de probabilidad para la verosimilitud y una para la a priori.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \propto P(X|C)P(C)$$

$$P(X|C) = P(x_1, \dots, x_n|C) = P(x_1|C) \cdot \dots \cdot P(x_n|C)$$

- Estimar los parámetros para estas distribuciones.

- Uso del clasificador:

- Clasificar un nuevo ejemplo a partir del modelo construido.

$$C = \max_{C \in \mathcal{C}} \arg \left\{ P(X|C)P(C) \right\}$$

- Bolsa de palabras: modelo para la representación de un documento en función de las palabras que contiene.

- Determinar vocabulario.

$$V = \{\text{azul}, \text{rojo}, \text{perro}, \text{gato}, \text{galleta}, \text{manzana}\}$$

- A partir de un texto.

*“el **perro azul** come una **galleta azul**”*

- Contar de acuerdo a un modelo.

Bernoulli: $d^B = \{1, 0, 1, 0, 1, 0\}$

Multinomial: $d^M = \{2, 0, 1, 0, 1, 0\}$

Ejemplo I: datos

- Tenemos un conjunto de entrenamiento de 11 documentos que pertenecen a las clases: Deportes (D) o Informática (I).

$$B_D = \begin{matrix} & w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

$$B_I = \begin{matrix} & w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{matrix}$$

- Vocabulario

$$V = \left| \begin{array}{llll} w_1 = \text{gol} & w_2 = \text{computación} & w_3 = \text{transmitir} & w_4 = \text{velocidad} \\ w_5 = \text{técnica} & w_6 = \text{defensa} & w_7 = \text{desempeño} & w_8 = \text{campo} \end{array} \right|$$

- Empleando un Clasificador Bayesiano Ingenuo clasificar los siguientes documentos:

$$x_1 = \{1, 0, 0, 1, 1, 1, 0, 1\} \quad x_2 = \{0, 1, 1, 0, 1, 0, 1, 0\}$$

Ejemplo I: construcción con EMV (A)

- Construcción del clasificador:
 - Asumir una distribución de probabilidad, en este caso Bernoulli.

$$P(D|C) = \prod_{t=1}^{|V|} Be(w_t; q) = \prod_{t=1}^{|V|} q^{w_t} (1 - q)^{1 - w_t}$$

- Estimar los parámetros de esta distribución usando el estimador de máxima verosimilitud.

$$\hat{q}_{(w_t|C)} = \frac{\sum_{i=1}^n x_i}{n} = \frac{n_C(w_t)}{N_C} \quad \hat{q}(C) = \frac{N_C}{N}$$

$n_C(w_t)$ = # de docs de clase C donde aparece w_t

N_C = # de docs de clase C

N = # de docs totales

- Uso del clasificador:

$$C = \max_{C \in \mathcal{C}} \arg \left\{ P(D|C)P(C) \right\}$$

$$C = \max_{C \in \{D, I\}} \arg \left\{ P(C) \prod_{t=1}^{|V|} (P_{(w_j|C)})^{w_j} (1 - P_{(w_j|C)})^{1-w_j} \right\}$$

Ejemplo I: estimación de parámetros con EMV

- Parámetros para la clase de deportes:

$$\begin{array}{llll}\hat{q}_{(w_1|D)} = \frac{1}{2} & \hat{q}_{(w_3|D)} = \frac{1}{3} & \hat{q}_{(w_5|D)} = \frac{1}{2} & \hat{q}_{(w_7|D)} = \frac{2}{3} \\ \hat{q}_{(w_2|D)} = \frac{1}{6} & \hat{q}_{(w_4|D)} = \frac{1}{2} & \hat{q}_{(w_6|D)} = \frac{2}{3} & \hat{q}_{(w_8|D)} = \frac{2}{3}\end{array}$$

- Parámetros para la clase de informática:

$$\begin{array}{llll}\hat{q}_{(w_1|I)} = \frac{1}{5} & \hat{q}_{(w_3|I)} = \frac{3}{5} & \hat{q}_{(w_5|I)} = \frac{1}{5} & \hat{q}_{(w_7|I)} = \frac{3}{5} \\ \hat{q}_{(w_2|I)} = \frac{3}{5} & \hat{q}_{(w_4|I)} = \frac{1}{5} & \hat{q}_{(w_6|I)} = \frac{1}{5} & \hat{q}_{(w_8|I)} = \frac{1}{5}\end{array}$$

- Parámetros de las clases:

$$\hat{q}_D = \frac{6}{11} \quad \hat{q}_I = \frac{5}{11}$$

Ejemplo I: clasificación con EMV (A)

$$P(C|d) \propto P(C)P(d|C) = P(C) \prod_{j=1}^{|w|} (P_{(w_j|C)})^{w_j} (1 - P_{(w_j|C)})^{1-w_j}$$

- $d_1 = \{1, 0, 0, 1, 1, 1, 0, 1\}$

$$P(D|d_1) \propto$$

$$P(I|d_1) \propto$$

- $d_2 = \{0, 1, 1, 0, 1, 0, 1, 0\}$

$$P(D|d_2) \propto$$

$$P(I|d_2) \propto$$

$$P(C|d) \propto P(C)P(d|C) = P(C) \prod_{j=1}^{|w|} (P_{(w_j|C)})^{w_j} (1 - P_{(w_j|C)})^{1-w_j}$$

- $d_1 = \{1, 0, 0, 1, 1, 1, 0, 1\}$

$$P(D|d_1) \propto \frac{6}{11} \left(\frac{1}{2} \times \frac{5}{6} \times \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} \times \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3} \right) = \frac{5}{891} = 5,6 \times 10^{-3}$$

$$P(I|d_1) \propto \frac{5}{11} \left(\frac{1}{5} \times \frac{2}{5} \times \frac{2}{5} \times \frac{1}{5} \times \frac{1}{5} \times \frac{1}{5} \times \frac{2}{5} \times \frac{1}{5} \right) = \frac{8}{859375} = 9,6 \times 10^{-6}$$

- d_1 se clasifica como un documento de deportes.

$$P(C|d) \propto P(C)P(d|C) = P(C) \prod_{j=1}^{|w|} (P_{(w_j|C)})^{w_j} (1 - P_{(w_j|C)})^{1-w_j}$$

- $d_2 = \{0, 1, 1, 0, 1, 0, 1, 0\}$

$$P(D|d_2) \propto \frac{6}{11} \left(\frac{1}{2} \times \frac{1}{6} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \right) = \frac{12}{14256} = 8,4 \times 10^{-4}$$

$$P(I|d_2) \propto \frac{5}{11} \left(\frac{4}{5} \times \frac{3}{5} \times \frac{3}{5} \times \frac{4}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{4}{5} \right) = \frac{34560}{4296875} = 8,0 \times 10^{-3}$$

- d_1 se clasifica como un documento de informática.

Ejemplo II: construcción con MAP

- Construcción del clasificador:

- Asumir una distribución de probabilidad, en este caso Bernoulli.

$$P(D|C) = \prod_{t=1}^{|V|} Be(w_t; q) = \prod_{t=1}^{|V|} q^{w_t} (1 - q)^{1 - w_t}$$

- Estimar los parámetros de esta distribución usando el estimador de máximo a posteriori.

$$\hat{q}_{MAP} = \frac{\sum_{i=1}^n x_i + \alpha - 1}{n + \beta + \alpha - 2}$$

- Empleando los equivalentes de suavizado de Laplace (add-one):
 - Parámetros de los atributos $\alpha = 2$ y $\beta = |V|$.
 - Parámetros de la clase $\alpha = 2$ y $\beta = |C|$.

Ejemplo II: parámetros con MAP

- Parámetros para la clase de deportes:

$$\begin{array}{llll}\hat{q}_{(w_1|D)} = \frac{4}{14} & \hat{q}_{(w_3|D)} = \frac{3}{14} & \hat{q}_{(w_5|D)} = \frac{2}{7} & \hat{q}_{(w_7|D)} = \frac{5}{14} \\ \hat{q}_{(w_2|D)} = \frac{1}{7} & \hat{q}_{(w_4|D)} = \frac{2}{7} & \hat{q}_{(w_6|D)} = \frac{5}{14} & \hat{q}_{(w_8|D)} = \frac{5}{14}\end{array}$$

- Parámetros para la clase de informática:

$$\begin{array}{llll}\hat{q}_{(w_1|I)} = \frac{2}{13} & \hat{q}_{(w_3|I)} = \frac{4}{13} & \hat{q}_{(w_5|I)} = \frac{2}{13} & \hat{q}_{(w_7|I)} = \frac{4}{13} \\ \hat{q}_{(w_2|I)} = \frac{4}{13} & \hat{q}_{(w_4|I)} = \frac{2}{13} & \hat{q}_{(w_6|I)} = \frac{2}{13} & \hat{q}_{(w_8|I)} = \frac{2}{13}\end{array}$$

- Parámetros de las clases:

$$\hat{q}_D = \frac{7}{13} \quad \hat{q}_I = \frac{6}{13}$$

Ejemplo II: clasificación con MAP (A)

$$P(C|d) \propto P(C)P(d|C) = P(C) \prod_{j=1}^{|w|} (P_{(w_j|C)})^{w_j} (1 - P_{(w_j|C)})^{1-w_j}$$

- $d_1 = \{1, 0, 0, 1, 1, 1, 0, 1\}$

$$P(D|d_1) \propto \frac{7}{13} \left(\frac{2}{7} \times \frac{6}{7} \times \frac{11}{14} \times \frac{2}{7} \times \frac{2}{7} \times \frac{5}{14} \times \frac{9}{14} \times \frac{5}{14} \right) = 6,9 \times 10^{-4}$$

$$P(I|d_1) \propto \frac{6}{13} \left(\frac{2}{13} \times \frac{9}{13} \times \frac{9}{13} \times \frac{2}{13} \times \frac{2}{13} \times \frac{2}{13} \times \frac{9}{13} \times \frac{2}{13} \right) = 1,13 \times 10^{-5}$$

- d_1 se clasifica como un documento de deportes.

Ejemplo II: clasificación con MAP (A)

$$P(C|d) \propto P(C)P(d|C) = P(C) \prod_{j=1}^{|w|} (P_{(w_j|C)})^{w_j} (1 - P_{(w_j|C)})^{1-w_j}$$

- $d_2 = \{0, 1, 1, 0, 1, 0, 1, 0\}$

$$P(D|d_2) \propto \frac{7}{13} \left(\frac{5}{7} \times \frac{1}{7} \times \frac{3}{14} \times \frac{5}{7} \times \frac{2}{7} \times \frac{9}{14} \times \frac{5}{14} \times \frac{9}{14} \right) = 3,54 \times 10^{-4}$$

$$P(I|d_2) \propto \frac{6}{13} \left(\frac{11}{13} \times \frac{4}{13} \times \frac{4}{13} \times \frac{11}{13} \times \frac{2}{13} \times \frac{11}{13} \times \frac{4}{13} \times \frac{11}{13} \right) = 1,06 \times 10^{-3}$$

- d_2 se clasifica como un documento de informática.