POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN

Universidad Nacional Autónoma de México

Aprendizaje automático.

Clasificador Bayesiano Ingenuo.

Ayudantes: Berenice y Ricardo

Febrero, 2020

Clasificador Bayesiano Ingenuo

- 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 y uso

- Construcción del clasificador:
  - Asumir para cada atributo una distribución de probabilidad para la verosimilitud y una para la a priori.

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

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

  - Estimar los parámetros para estas distribuciones.

- Uso del clasificador:
  - Clasificar un nuevo ejemplo a partir del modelo construido.

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

Ejemplo I: datos

- Una estación de radio clasificará su audiencia en jóvenes y adultos a partir de sus gustos musicales. Se realizó una encuesta que consistía en indicar si les gustaba cierta agrupación (1) o no (0).

Pink Floyd	**The Beatles**	**R.E.M.**	**Nirvana**	**Queen**	**Oasis**	**Joven (J) / Adulto (A)**
1	0	0	1	1	1	J
1	1	0	1	1	0	J
1	1	1	0	0	1	J
1	0	1	0	0	1	J
1	0	0	0	1	0	J
1	1	1	0	0	0	J
1	1	0	0	1	1	A
1	1	1	0	0	1	A
1	1	1	1	1	0	A
1	1	1	0	1	0	A
1	1	1	0	1	1	A
1	1	0	1	1	0	A
1	1	0	1	0	0	A

Ejemplo I: nuevos datos

- Entrena un clasificador bayesiano ingenuo y clasifica los siguientes vectores:
  - $X^{(1)} = (1, 1, 0, 1, 1, 0)$
  - $X^{(2)} = (1, 0, 1, 1, 1, 1)$
  - $X^{(3)} = (1, 1, 0, 0, 0, 0)$
  - $X^{(4)} = (1, 1, 1, 1, 1, 1)$
  - $X^{(5)} = (0, 1, 1, 1, 1, 1)$

Ejemplo I: construcción con EMV (A)

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

$$P(X^{(i)}|C) = \prod_{j=1}^{n} Ber(X_j^{(i)}; q_j) = \prod_{j=1}^{n} q_j^{X_j^{(i)}} (1-q)^{1-X_j^{(i)}}$$

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

$$\hat{q}_{(X_j^{(i)}|C)} = \frac{\sum_{i=1}^{m} X_j^{(i)}}{m_c} = \frac{\text{\# de entrevistados de la clase } C \text{ que les gusta } X_j}{\text{\# de entrevistados de la clase } C}$$

$$\hat{q}_{(C)} = \frac{m_c}{m} = \frac{\text{\# de entrevistados de la clase } C}{\text{número total de entrevistados}}$$

Ejemplo I: construcción con EMV (B)

- Uso del clasificador:

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

$$C = \max_{C \in \{A,J\}} \arg \left\{ P(C) \prod_{j=1}^{n} (P_{(X_j^{(i)}|C)})^{X_j^{(i)}} (1 - P_{(X_j^{(i)}|C)})^{1-X_j^{(i)}} \right\}$$

Ejemplo I: parámetros con EMV

- Parámetros para la clase de joven (J):

$$\hat{q}_{(PinkFloyd|J)} = 1 \qquad \hat{q}_{(R.E.M|J)} = \frac{1}{2} \qquad \hat{q}_{(Queen|J)} = \frac{1}{2}$$

$$\hat{q}_{(Beatles|J)} = \frac{1}{2} \qquad \hat{q}_{(Nirvana|J)} = \frac{1}{3} \qquad \hat{q}_{(Oasis|J)} = \frac{1}{2}$$

- Parámetros para la clase de adulto:

$$\hat{q}_{(PinkFloyd|A)} = 1 \qquad \hat{q}_{(R.E.M|A)} = \frac{4}{7} \qquad \hat{q}_{(Queen|A)} = \frac{5}{7}$$

$$\hat{q}_{(Beatles|A)} = 1 \qquad \hat{q}_{(Nirvana|A)} = \frac{3}{7} \qquad \hat{q}_{(Oasis|A)} = \frac{3}{7}$$

- Parámetros de las clases:

$$\hat{q}_J = \frac{6}{13} \qquad \hat{q}_A = \frac{7}{13}$$

Ejemplo I: clasificación con EMV (A)

$$P(C|X^{(i)}) \propto P(C)P(X^{(i)}|C) = P(C)\prod_{j=1}^{n}(P_{(X_j^{(i)}|C)})^{X_j^{(i)}}(1 - P_{(X_j^{(i)}|C)})^{1-X_j^{(i)}}$$

- $X^{(1)} = (1, 1, 0, 1, 1, 0)$

  $P(J|X^{(1)}) \propto$

  $P(A|X^{(1)}) \propto$

- $X^{(2)} = (1, 0, 1, 1, 1, 1)$

  $P(J|X^{(2)}) \propto$

  $P(A|X^{(2)}) \propto$

Ejemplo I: clasificación con EMV (B)

- $X^{(1)} = (1, 1, 0, 1, 1, 0)$

$$P(J|X^{(1)}) \propto \frac{6}{13}\left(1 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2}\right) = \frac{1}{104} = 9.61 \times 10^{-3}$$

$$P(A|X^{(1)}) \propto \frac{7}{13}\left(1 \times 1 \times \frac{3}{7} \times \frac{3}{7} \times \frac{5}{7} \times \frac{4}{7}\right) = \frac{180}{4459} = .0403$$

dado que $P(A|X^{(1)}) > P(J|X^{(1)})$ entonces $X^{(1)}$ se clasifica como Adulto.

- $X^{(2)} = (1, 0, 1, 1, 1, 1)$

$$P(J|X^{(2)}) \propto \frac{6}{13}\left(1 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2}\right) = \frac{1}{104} = 9.61 \times 10^{-3}$$

$$P(A|X^{(2)}) \propto \frac{7}{13}\left(1 \times 0 \times \frac{4}{7} \times \frac{3}{7} \times \frac{5}{7} \times \frac{3}{7}\right) = 0$$

dado que $P(J|X^{(2)}) > P(A|X^{(2)})$ entonces $X^{(2)}$ se clasifica como Joven.

Ejemplo I: clasificación con EMV (C)

$$P(C|X^{(i)}) \propto P(C)P(X^{(i)}|C) = P(C)\prod_{j=1}^{n}(P_{(X_j^{(i)}|C)})^{X_j^{(i)}}(1 - P_{(X_j^{(i)}|C)})^{1-X_j^{(i)}}$$

- $X^{(3)} = (1, 1, 0, 0, 0, 0)$

  $P(J|X^{(3)}) \propto$

  $P(A|X^{(3)}) \propto$

- $X^{(4)} = (1, 1, 1, 1, 1, 1)$

  $P(J|X^{(4)}) \propto$

  $P(A|X^{(4)}) \propto$

Ejemplo I: clasificación con EMV (D)

- $X^{(3)} = (1, 1, 0, 0, 0, 0)$

$$P(J|X^{(3)}) \propto \frac{6}{13}\left(1 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2}\right) = \frac{1}{52} = .019$$

$$P(A|X^{(3)}) \propto \frac{7}{13}\left(1 \times 1 \times \frac{3}{7} \times \frac{4}{7} \times \frac{2}{7} \times \frac{4}{7}\right) = \frac{96}{4459} = .021$$

dado que $P(A|X^{(3)}) > P(J|X^{(3)})$ entonces $X^{(3)}$ se clasifica como Adulto.

- $X^{(4)} = (1, 1, 1, 1, 1, 1)$

$$P(J|X^{(4)}) \propto \frac{6}{13}\left(1 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2}\right) = \frac{1}{104} = 9.61 \times 10^{-3}$$

$$P(A|X^{(4)}) \propto \frac{7}{13}\left(1 \times 1 \times \frac{4}{7} \times \frac{3}{7} \times \frac{5}{7} \times \frac{3}{7}\right) = \frac{180}{4459} = 0.040$$

dado que $P(A|X^{(4)}) > P(J|X^{(4)})$ entonces $X^{(4)}$ se clasifica como Adulto.

Ejemplo I: clasificación con EMV (E)

$$P(C|X^{(i)}) \propto P(C)P(X^{(i)}|C) = P(C)\prod_{j=1}^{n}(P_{(X_j^{(i)}|C)})^{X_j^{(i)}}(1 - P_{(X_j^{(i)}|C)})^{1-X_j^{(i)}}$$

- $X^{(5)} = (0, 1, 1, 1, 1, 1)$

  $P(J|X^{(5)}) \propto$

  $P(A|X^{(5)}) \propto$

Ejemplo I: clasificación con EMV (F)

- $X^{(5)} = (0, 1, 1, 1, 1, 1)$

$$P(J|X^{(5)}) \propto \frac{6}{13}\left(0 \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2}\right) = 0$$

$$P(A|X^{(5)}) \propto \frac{7}{13}\left(0 \times 1 \times \frac{4}{7} \times \frac{3}{7} \times \frac{5}{7} \times \frac{3}{7}\right) = 0$$

$X^{(5)}$ no puede ser clasificado usando el estimador de máxima verosimilitud.

Ejemplo I: construcción con MAP

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

$$P(X^{(i)}|C) = \prod_{j=1}^{n} Ber(X_j^{(i)}; q_j) = \prod_{j=1}^{n} q_j^{X_j^{(i)}} (1-q)^{1-X_j^{(i)}}$$

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

$$\hat{q}_{(X_j^{(i)}|C)} = \frac{\sum_{i=1}^{m} X_j^{(i)} + \alpha - 1}{m_c + \beta + \alpha - 2}$$

$$\hat{q}_{(C)} = \frac{m_c + \alpha - 1}{m + \beta + \alpha - 2}$$

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

Ejemplo I: parámetros con MAP

- Parámetros para la clase de joven (J):

$$\hat{q}(PinkFloyd|J) = \frac{7}{12} \qquad \hat{q}(R.E.M|J) = \frac{1}{3} \qquad \hat{q}(Queen|J) = \frac{1}{3}$$
$$\hat{q}(Beatles|J) = \frac{1}{3} \qquad \hat{q}(Nirvana|J) = \frac{1}{4} \qquad \hat{q}(Oasis|J) = \frac{1}{3}$$

- Parámetros para la clase de adulto:

$$\hat{q}(PinkFloyd|A) = \frac{8}{13} \qquad \hat{q}(R.E.M|A) = \frac{5}{13} \qquad \hat{q}(Queen|A) = \frac{6}{13}$$
$$\hat{q}(Beatles|A) = \frac{8}{13} \qquad \hat{q}(Nirvana|A) = \frac{4}{13} \qquad \hat{q}(Oasis|A) = \frac{4}{13}$$

- Parámetros de las clases:

$$\hat{q}_J = \frac{7}{15} \qquad \hat{q}_A = \frac{8}{15}$$

Ejemplo I: clasificación con MAP (A)

- $X^{(1)} = (1, 1, 0, 1, 1, 0)$

  $$P(J|X^{(1)}) \propto \frac{7}{15}\left(\frac{7}{12} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{4} \times \frac{1}{3} \times \frac{2}{3}\right) = \frac{49}{14580} = 3.36 \times 10^{-3}$$

  $$P(A|X^{(1)}) \propto \frac{8}{15}\left(\frac{8}{13} \times \frac{8}{13} \times \frac{8}{13} \times \frac{9}{13} \times \frac{7}{13} \times \frac{9}{13}\right) = 0.012$$

  dado que $P(A|X^{(1)}) > P(J|X^{(1)})$ entonces $X^{(1)}$ se clasifica como Adulto.

- $X^{(2)} = (1, 0, 1, 1, 1, 1)$

  $$P(J|X^{(2)}) \propto \frac{7}{15}\left(\frac{7}{12} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{4} \times \frac{1}{3} \times \frac{1}{3}\right) = \frac{49}{29160} = 1.68 \times 10^{-3}$$

  $$P(A|X^{(2)}) \frac{8}{15}\left(\frac{8}{13} \times \frac{5}{13} \times \frac{5}{13} \times \frac{4}{13} \times \frac{6}{13} \times \frac{4}{13}\right) = 2.12 \times 10^{-3}$$

  dado que $P(A|X^{(2)}) > P(J|X^{(2)})$ entonces $X^{(2)}$ se clasifica como Adulto.

Ejemplo I: clasificación con MAP (B)

- $X^{(3)} = (1, 1, 0, 0, 0, 0)$

  $$P(J|X^{(3)}) \propto \frac{7}{15}\left(\frac{7}{12} \times \frac{1}{3} \times \frac{2}{3} \times \frac{3}{4} \times \frac{2}{3} \times \frac{2}{3}\right) = \frac{49}{2430} = 0.20$$

  $$P(A|X^{(3)}) \propto \frac{8}{15}\left(\frac{8}{13} \times \frac{8}{13} \times \frac{8}{13} \times \frac{9}{13} \times \frac{7}{13} \times \frac{9}{13}\right) = 0.032$$

  dado que $P(A|X^{(3)}) > P(J|X^{(3)})$ entonces $X^{(3)}$ se clasifica como Adulto.

- $X^{(4)} = (1, 1, 1, 1, 1, 1)$

  $$P(J|X^{(4)}) \propto \frac{7}{15}\left(\frac{7}{12} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{4} \times \frac{1}{3} \times \frac{1}{3}\right) = \frac{49}{58320} = 8.40 \times 10^{-4}$$

  $$P(A|X^{(4)}) \propto \frac{8}{15}\left(\frac{8}{13} \times \frac{8}{13} \times \frac{5}{13} \times \frac{4}{13} \times \frac{6}{13} \times \frac{4}{13}\right) = 3.39 \times 10^{-3}$$

  dado que $P(A|X^{(4)}) > P(J|X^{(4)})$ entonces $X^{(4)}$ se clasifica como Adulto.

Ejemplo I: clasificación con MAP (C)

- $X^{(5)} = (0, 1, 1, 1, 1, 1)$

$$P(J|X^{(5)}) \propto \frac{7}{15}\left(\frac{5}{12} \times \frac{1}{3} \times \frac{1}{3} \times \frac{1}{4} \times \frac{1}{3} \times \frac{1}{3}\right) = \frac{7}{11664} = .0006$$

$$P(A|X^{(5)}) \propto \frac{8}{15}\left(\frac{5}{13} \times \frac{8}{13} \times \frac{5}{13} \times \frac{4}{13} \times \frac{6}{13} \times \frac{4}{13}\right) = .0021$$

dado que $P(A|X^{(5)}) > P(J|X^{(5)})$ entonces $X^{(5)}$ se clasifica como Adulto.

Clasificación de texto

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

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

  - A partir de un texto.

"el perro azul come una galleta azul"

  - Contar de acuerdo a un modelo.

$$\begin{aligned} \text{Bernoulli:} \quad & d^B = \{1, 0, 1, 0, 1, 0\} \\ \text{Multinomial:} \quad & d^M = \{2, 0, 1, 0, 1, 0\} \end{aligned}$$

Ejemplo II: datos

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

$$B_D = \begin{array}{c} \begin{array}{cccccccc} w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array} \qquad B_I = \begin{array}{c} \begin{array}{cccccccc} w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 & w_8 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

- Vocabulario

$$V = \begin{vmatrix} w_1 = \textit{gol} & w_2 = \textit{computacin} & w_3 = \textit{transmitir} & w_4 = \textit{velocidad} \\ w_5 = \textit{técnica} & w_6 = \textit{defensa} & w_7 = \textit{desempeño} & w_8 = \textit{campo} \end{vmatrix}$$

- Empleando un Clasificador Bayesiano Ingenuo clasificar los siguientes documentos:

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

Ejemplo II: 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.

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

$$n_C(w_t) = \#\text{ de docs de clase } C \text{ donde aparece } w_t$$

$$N_C = \#\text{ de docs de clase } C$$

$$N = \#\text{ de docs totales}$$

Ejemplo II: construcción con EMV (B)

- 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 II: parámetros con EMV

- Parámetros para la clase de deportes:

$$\hat{P}_{(w_1|D)} = \frac{1}{2} \qquad \hat{P}_{(w_3|D)} = \frac{1}{3} \qquad \hat{P}_{(w_5|D)} = \frac{1}{2} \qquad \hat{P}_{(w_7|D)} = \frac{2}{3}$$

$$\hat{P}_{(w_2|D)} = \frac{1}{6} \qquad \hat{P}_{(w_4|D)} = \frac{1}{2} \qquad \hat{P}_{(w_6|D)} = \frac{2}{3} \qquad \hat{P}_{(w_8|D)} = \frac{2}{3}$$

- Parámetros para la clase de informática:

$$\hat{P}_{(w_1|I)} = \frac{1}{5} \qquad \hat{P}_{(w_3|I)} = \frac{3}{5} \qquad \hat{P}_{(w_5|I)} = \frac{1}{5} \qquad \hat{P}_{(w_7|I)} = \frac{3}{5}$$

$$\hat{P}_{(w_2|I)} = \frac{3}{5} \qquad \hat{P}_{(w_4|I)} = \frac{1}{5} \qquad \hat{P}_{(w_6|I)} = \frac{1}{5} \qquad \hat{P}_{(w_8|I)} = \frac{1}{5}$$

- Parámetros de las clases:

$$\hat{P}_D = \frac{6}{11} \qquad \hat{P}_I = \frac{5}{11}$$

Ejemplo II: clasificación con EMV (A)

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

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

  $P(D|d_1) \propto$

  $P(I|d_1) \propto$

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

  $P(D|d_2) \propto$

  $P(I|d_2) \propto$

Ejemplo II: clasificación con EMV (B)

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

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

$$P(D|d_1) \propto \frac{6}{11}\left(\frac{1}{2}\times\frac{5}{6}\times\frac{2}{3}\times\frac{1}{2}\times\frac{1}{2}\times\frac{2}{3}\times\frac{1}{3}\times\frac{2}{3}\right) = \frac{5}{891} = 5.6\times 10^{-3}$$

$$P(I|d_1) \propto \frac{5}{11}\left(\frac{1}{5}\times\frac{2}{5}\times\frac{2}{5}\times\frac{1}{5}\times\frac{1}{5}\times\frac{1}{5}\times\frac{2}{5}\times\frac{1}{5}\right) = \frac{8}{859375} = 9.6\times 10^{-6}$$

- d_1 se clasifica como un documento de deportes.

Ejemplo II: clasificación con EMV (C)

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

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

$$P(D|d_2) \propto \frac{6}{11}\left(\frac{1}{2}\times\frac{1}{6}\times\frac{1}{3}\times\frac{1}{2}\times\frac{1}{2}\times\frac{1}{3}\times\frac{2}{3}\times\frac{1}{3}\right) = \frac{12}{14256} = 8.4\times 10^{-4}$$

$$P(I|d_2) \propto \frac{5}{11}\left(\frac{4}{5}\times\frac{3}{5}\times\frac{3}{5}\times\frac{4}{5}\times\frac{1}{5}\times\frac{4}{5}\times\frac{3}{5}\times\frac{4}{5}\right) = \frac{34560}{4296875} = 8.0\times 10^{-3}$$

- 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.

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

  - 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: clasificación con MAP (A)

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

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

  $P(D|d_1) \propto$

  $P(I|d_1) \propto$

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

  $P(D|d_2) \propto$

  $P(I|d_2) \propto$