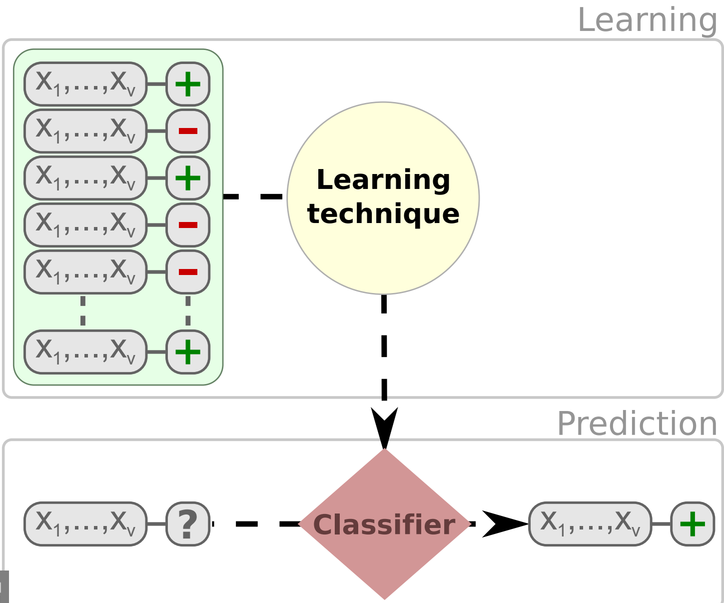
Aprendizaje no supervisado

VC08: Aprendizaje semi-supervisado

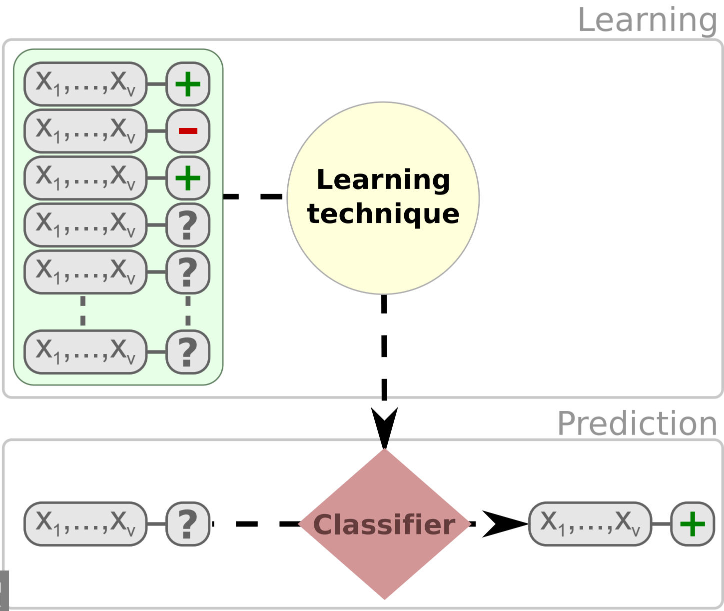
Félix José Fuentes Hurtado
felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia

Aprendizaje semi-supervisado

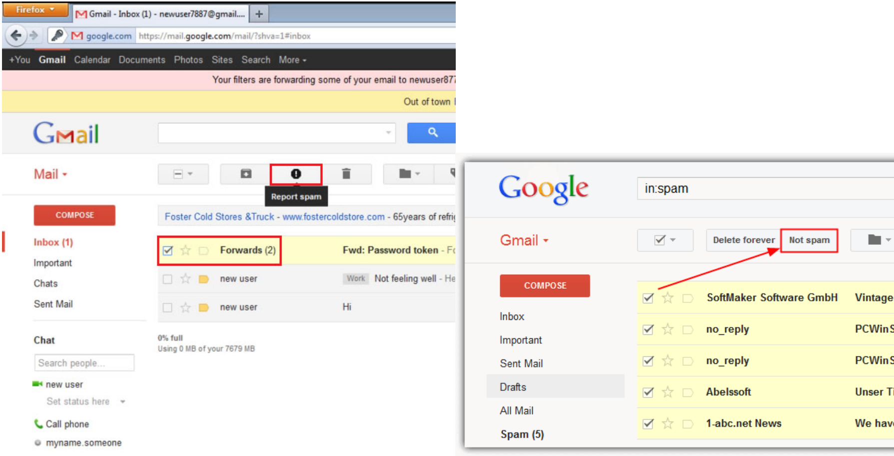


Aprendizaje semi-supervisado



Aprendizaje semi-supervisado

Aprendizaje semi-supervisado

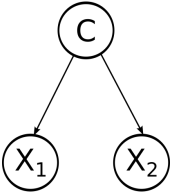


EM

Estrategia para estimar los parámetros de máxima verosimilitud (MLE) cuando hay datos incompletos.

¿Por qué no se pueden obtener directamente?

X_1	X_2	C
1	0	0
0	0	?
0	1	?
0	1	?
1	0	1
0	0	1
1	1	?
1	1	?
0	1	0
1	1	?



C	$p(C)$
0	$?/10$
1	$?/10$

EM

Algoritmo *Expectation-Maximization*:

Procedimiento iterativo de dos pasos (E-M) que permite obtener los parámetros de máxima verosimilitud cuando hay datos perdidos (valores perdidos, variables latentes, etc.)

E-step: Se estima el valor de los datos perdidos usando la esperanza condicional de la verosimilitud

M-step: Se estiman unos nuevos parámetros dados los datos completados en el paso E.

Convergencia:

- Máximo (local)
- Casos raros: Punto de silla

EM

Algoritmo *Expectation-Maximization*:

Procedimiento iterativo de dos pasos (E-M) que permite obtener los parámetros de máxima verosimilitud cuando hay datos perdidos (valores perdidos, variables latentes, etc.)

E-step:

$$Q(\theta;\theta^t) = E_{Z|X,\theta^t}\left[\log L(\theta;X,Z)\right]$$

M-step: Choose θ^{t+1} such that, for all $\theta \in \Theta$:

$$Q(\theta^{t+1};\theta^t) \geq Q(\theta;\theta^t)$$

Donde Z son los datos perdidos, X los observados, y θ los parámetros del modelo. Se define verosimilitud como:

$$L(\theta;X,Z) = p(X,Z;\theta)$$

EM

Algoritmo *Expectation-Maximization*:

Procedimiento iterativo de dos pasos (E-M) que permite obtener los parámetros de máxima verosimilitud cuando hay datos perdidos (valores perdidos, variables latentes, etc.)

E-step:

$$Q(\theta;\theta^t) = E_{Z|X,\theta^t}\left[\log L(\theta;X,Z)\right]$$

M-step:

$$\theta^{t+1} = \operatorname*{argmáx}_{\theta} Q(\theta;\theta^t)$$

Donde Z son los datos perdidos, X los observados, y θ los parámetros del modelo. Se define verosimilitud como:

$$L(\theta;X,Z) = p(X,Z;\theta)$$

EM

Verosimilitud:

$$L(\theta; X, Z) = p(X, Z; \theta)$$

$$L(\theta; X) = p(X; \theta) = \sum_{Z} p(X, Z; \theta)$$

$$p(X; \theta) = \frac{p(X, Z; \theta)}{p(Z|X; \theta)}$$

¿Maximizando Q maximizamos la verosimilitud?

$$\log p(X; \theta) = \log p(X, Z; \theta) - \log p(Z|X; \theta)$$

EM

Verosimilitud:

$$L(\theta; X, Z) = p(X, Z; \theta)$$

$$L(\theta; X) = p(X; \theta) = \sum_Z p(X, Z; \theta)$$

$$p(X; \theta) = \frac{p(X, Z; \theta)}{p(Z|X; \theta)}$$

¿Maximizando Q maximizamos la verosimilitud?

$$\log p(X; \theta) = \sum_Z p(Z|X; \theta^t) \log p(X, Z; \theta) - \sum_Z p(Z|X; \theta^t) \log p(Z|X; \theta)$$

EM

Verosimilitud:

$$L(\theta; X, Z) = p(X, Z; \theta)$$

$$L(\theta; X) = p(X; \theta) = \sum_Z p(X, Z; \theta)$$

$$p(X; \theta) = \frac{p(X, Z; \theta)}{p(Z|X; \theta)}$$

¿Maximizando Q maximizamos la verosimilitud?

$$\log p(X|\theta) = Q(\theta; \theta^t) + H(\theta; \theta^t)$$

EM

Verosimilitud:

$$L(\theta; X, Z) = p(X, Z; \theta)$$

$$L(\theta; X) = p(X; \theta) = \sum_{Z} p(X, Z; \theta)$$

$$p(X; \theta) = \frac{p(X, Z; \theta)}{p(Z|X; \theta)}$$

¿Maximizando Q maximizamos la verosimilitud?

$$\log p(X|\theta) = Q(\theta; \theta^t) + H(\theta; \theta^t)$$

$$\log p(X|\theta) - \log p(X|\theta^t) = Q(\theta; \theta^t) - Q(\theta^t; \theta^t) + H(\theta; \theta^t) - H(\theta^t; \theta^t)$$

EM

Verosimilitud:

$$L(\theta; X, Z) = p(X, Z; \theta)$$

$$L(\theta; X) = p(X; \theta) = \sum_{Z} p(X, Z; \theta)$$

$$p(X; \theta) = \frac{p(X, Z; \theta)}{p(Z|X; \theta)}$$

¿Maximizando Q maximizamos la verosimilitud?

$$\log p(X|\theta) = Q(\theta; \theta^t) + H(\theta; \theta^t)$$

$$\log p(X|\theta) - \log p(X|\theta^t) = Q(\theta; \theta^t) - Q(\theta^t; \theta^t) + C$$

con $C \geq 0$.

EM

Verosimilitud:

$$L(\theta; X, Z) = p(X, Z; \theta)$$

$$L(\theta; X) = p(X; \theta) = \sum_{Z} p(X, Z; \theta)$$

$$p(X; \theta) = \frac{p(X, Z; \theta)}{p(Z|X; \theta)}$$

¿Maximizando Q maximizamos la verosimilitud?

$$\log p(X|\theta) = Q(\theta; \theta^t) + H(\theta; \theta^t)$$

$$\log p(X|\theta) - \log p(X|\theta^t) \geq Q(\theta; \theta^t) - Q(\theta^t; \theta^t)$$

EM en la práctica

Aprendizaje del modelo (NB)

Paso E:

Determinista

X_1	X_2	C
1	0	0
0	0	?
0	1	?
0	1	?
1	0	1
0	0	1
1	1	?
1	0	?
0	1	0
1	1	?

Probabilista

X_1	X_2	C
1	0	0
0	0	?
0	1	?
0	1	?
1	0	1
0	0	1
1	1	?
1	0	?
0	1	0
1	1	?

EM en la práctica

Aprendizaje del modelo (NB)

Paso E:

Determinista

X_1	X_2	C
1	0	0
0	0	1
0	1	0
0	1	1
1	0	1
0	0	1
1	1	0
1	0	1
0	1	0
1	1	0

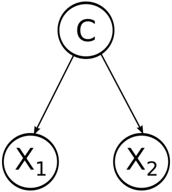
Probabilista

X_1	X_2	C	
		0	1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.7	0.3
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.8	0.2
1	0	0.3	0.7
0	1	1.0	0.0
1	1	0.6	0.4

EM

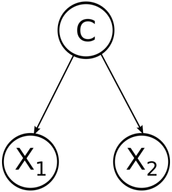
Aprendizaje del modelo

X_1	X_2	C
1	0	0
0	0	?
0	1	?
0	1	?
1	0	1
0	0	1
1	1	?
1	0	?
0	1	0
1	1	?



C	$p(C)$
0	?/10
1	?/10

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5

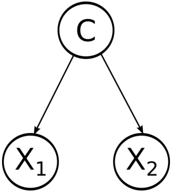


C	$p(C)$
0	
1	

EM

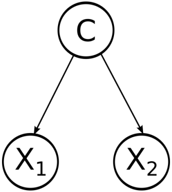
Aprendizaje del modelo

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5



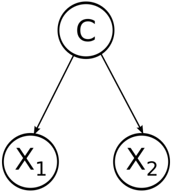
C	$p(C)$
0	$5/10 = 0{,}5$
1	$5/10 = 0{,}5$

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5



X_1	C	$p(X_1 \mid C)$
0	0	
1	0	
0	1	
1	1	

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5

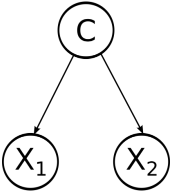


X_1	C	$p(X_1\|C)$
0	0	$2{,}5/5 = 0{,}50$
1	0	$2{,}5/5 = 0{,}50$
0	1	$2{,}5/5 = 0{,}50$
1	1	$2{,}5/5 = 0{,}50$

EM

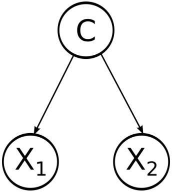
Aprendizaje del modelo

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5



X_2	C	$p(X_2 \mid C)$
0	0	
1	0	
0	1	
1	1	

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5

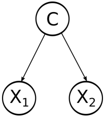


X_2	C	$p(X_2 \mid C)$
0	0	$2/5 = 0{,}40$
1	0	$3/5 = 0{,}60$
0	1	$3/5 = 0{,}60$
1	1	$2/5 = 0{,}40$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5

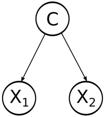


C	$p(C)$
0	0.5
1	0.5

X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

$$\hat{c} = \operatorname*{argmáx}_{c} p(c) \prod_{i=1}^{2} p(x_i|c)$$

X_1	X_2	C	
		0	1
1	0	1.0	0.0
0	0	0.5	0.5
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5



C	$p(C)$
0	0.5
1	0.5

X_i	C	$p(X_1 \mid C)$	$p(X_2 \mid C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

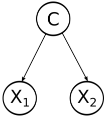
$$p(c = 0 \mid \mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i \mid 0)$$

$$p(c = 1 \mid \mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i \mid 1)$$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.10	0.15
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5



C	$p(C)$
0	0.5
1	0.5

X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

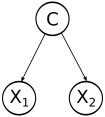
$$p(c=0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

$$p(c=1|\mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i|1)$$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.5	0.5
0	1	0.5	0.5
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.5	0.5
1	0	0.5	0.5
0	1	1.0	0.0
1	1	0.5	0.5



C	$p(C)$
0	0.5
1	0.5

X_i	C	$p(X_1 \mid C)$	$p(X_2 \mid C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

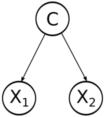
$$p(c = 0 \mid \mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i \mid 0)$$

$$p(c = 1 \mid \mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i \mid 1)$$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.15	0.10
0	1	0.15	0.10
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.10	0.15
1	0	0.15	0.10
0	1	1.0	0.0
1	1	0.15	0.10



C	$p(C)$
0	0.5
1	0.5

X_i	C	$p(X_1 \mid C)$	$p(X_2 \mid C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

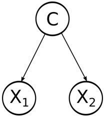
$$p(c = 0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

$$p(c = 1|\mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i|1)$$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4



C	$p(C)$
0	0.5
1	0.5

X_i	C	$p(X_1 \vert C)$	$p(X_2 \vert C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

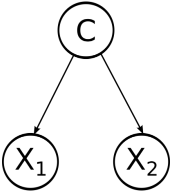
$$p(c = 0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

$$p(c = 1|\mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i|1)$$

EM

Re-aprender el modelo

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4

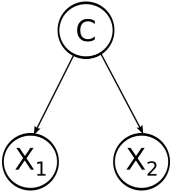


C	$p(C)$
0	
1	

EM

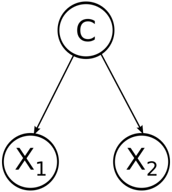
Re-aprender el modelo

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4



C	$p(C)$
0	0,52
1	0,48

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4

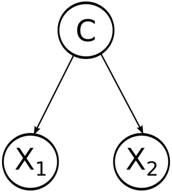


X_1	C	$p(X_1 \mid C)$
0	0	
1	0	
0	1	
1	1	

EM

Re-aprender el modelo

X_1	X_2	C	
		0	1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4

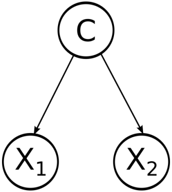


X_1	C	$p(X_1 \mid C)$
0	0	$2{,}6/5{,}2 = 0{,}5$
1	0	$2{,}6/5{,}2 = 0{,}5$
0	1	$2{,}4/4{,}8 = 0{,}5$
1	1	$2{,}4/4{,}8 = 0{,}5$

EM

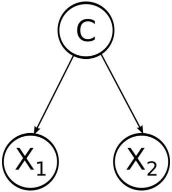
Re-aprender el modelo

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4



X_2	C	$p(X_2 \mid C)$
0	0	
1	0	
0	1	
1	1	

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4

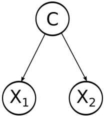


X_2	C	$p(X_2\|C)$
0	0	$2{,}0/5{,}2 = 0{,}385$
1	0	$3{,}2/5{,}2 = 0{,}615$
0	1	$3{,}0/4{,}8 = 0{,}625$
1	1	$1{,}8/4{,}8 = 0{,}375$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4



C	$p(C)$
0	0.52
1	0.48

X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.385
1	0	0.5	0.615
0	1	0.5	0.625
1	1	0.5	0.375

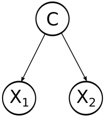
$$p(c=0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

$$p(c=1|\mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i|1)$$

EM

Re-estimación de pesos

X_1	X_2	$C = 0$	$C = 1$
1	0	1.0	0.0
0	0	0.41	0.59
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4



C	$p(C)$
0	0.52
1	0.48

X_i	C	$p(X_1 \mid C)$	$p(X_2 \mid C)$
0	0	0.5	0.385
1	0	0.5	0.615
0	1	0.5	0.625
1	1	0.5	0.375

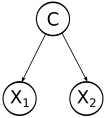
$$p(c = 0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

$$p(c = 1|\mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i|1)$$

EM

Re-estimación de pesos

X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.41	0.59
0	1	0.64	0.36
0	1	0.64	0.36
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.64	0.36
1	0	0.41	0.59
0	1	1.0	0.0
1	1	0.64	0.36



C	$p(C)$
0	0.52
1	0.48

X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.385
1	0	0.5	0.615
0	1	0.5	0.625
1	1	0.5	0.375

$$p(c=0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

$$p(c=1|\mathbf{x}) = \frac{1}{\theta} p(1) \prod_{i=1}^{2} p(x_i|1)$$

En la práctica...

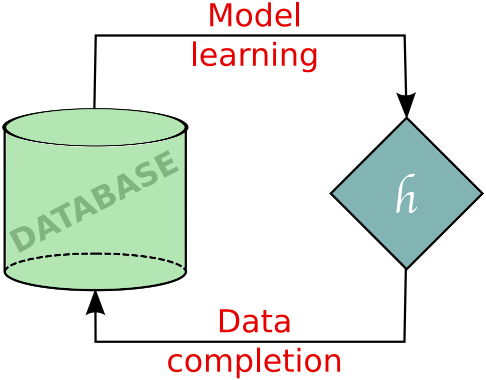
- Estimación de parámetros: Laplace Smoothing

$$\theta_i = \frac{N_i + 1}{N + |L|}$$

- Cálculo de probabilidades: cálculo logarítmico

$$\hat{c} = \underset{c}{\operatorname{argmáx}} \exp^{\left[\log p(c) + \sum_{i=1}^{2} \log p(x_i|c)\right]}$$

EM



Aprendizaje no supervisado

VC08: Aprendizaje semi-supervisado

Félix José Fuentes Hurtado
felixjose.fuentes@campusviu.es

Universidad Internacional de Valencia

Universidad Internacional de Valencia