Aprendiendo con incertidumbre
Expectation-Maximization algorithm

Jerónimo Hernández González

Exploración y análisis de datos

Máster Universitario en Ingeniería
Computacional y Sistemas Inteligentes

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

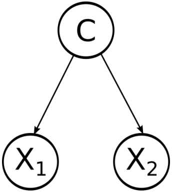
MDE
Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School

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:

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

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

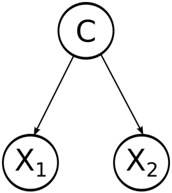
Probabilista

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	?

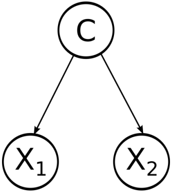


C	$p(C)$
0	$?/10$
1	$?/10$

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

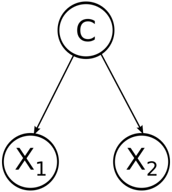


C	$p(C)$
0	
1	

EM

Aprendizaje del modelo

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

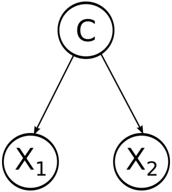


C	$p(C)$
0	$5/10 = 0{,}5$
1	$5/10 = 0{,}5$

EM

Aprendizaje del modelo

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

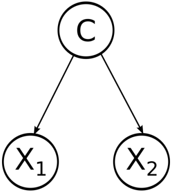


X_1	C	$p(X_1 \mid C)$
0	0	
1	0	
0	1	
1	1	

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

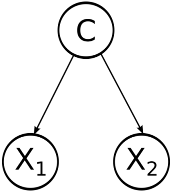


X_1	C	$p(X_1 \mid C)$
0	0	$2{,}5/5 = 0{,}50$
1	0	$2{,}5/5 = 0{,}50$
0	1	$2{,}5/5 = 0{,}50$
1	1	$2{,}5/5 = 0{,}50$

EM

Aprendizaje del modelo

X_1	X_2	C	
		0	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

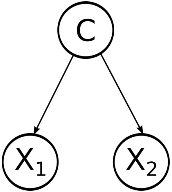


X_2	C	$p(X_2 \mid C)$
0	0	
1	0	
0	1	
1	1	

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

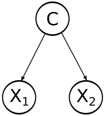


X_2	C	$p(X_2 \| C)$
0	0	$2/5 = 0{,}40$
1	0	$3/5 = 0{,}60$
0	1	$3/5 = 0{,}60$
1	1	$2/5 = 0{,}40$

EM

Re-estimación de pesos

X_1	X_2	C	
		0	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



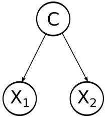
C	$p(C)$
0	0.5
1	0.5

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

$$\hat{c} = \operatorname*{argmax}_c p(c) \prod_{i=1}^{2} p(x_i \mid c)$$

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

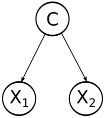
X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

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

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

EM

Re-estimación de pesos



X_1	X_2	$C=0$	$C=1$
1	0	1.0	0.0
0	0	0.10	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

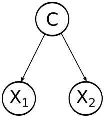
X_i	C	$p(X_1 \vert C)$	$p(X_2 \vert C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

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

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

EM

Re-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

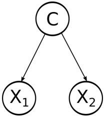
X_i	C	$p(X_1 \mid C)$	$p(X_2 \mid C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

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

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

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	0.15	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

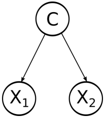
X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.4
1	0	0.5	0.6
0	1	0.5	0.6
1	1	0.5	0.4

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

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

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

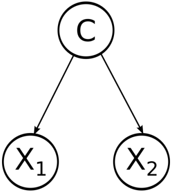
$$p(c=0|\mathbf{x}) = \frac{1}{\theta} p(0) \prod_{i=1}^{2} p(x_i|0)$$

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

EM

Re-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

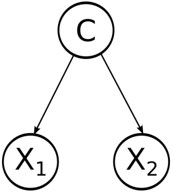


C	$p(C)$
0	
1	

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

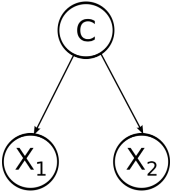


C	$p(C)$
0	0,52
1	0,48

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

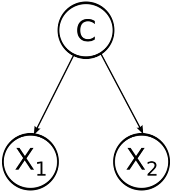


X_1	C	$p(X_1 \mid C)$
0	0	
1	0	
0	1	
1	1	

EM

Re-aprender el modelo

X_1	X_2	C = 0	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

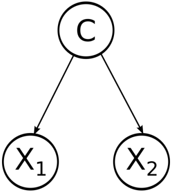


X_1	C	$p(X_1\|C)$
0	0	$2{,}6/5{,}2 = 0{,}5$
1	0	$2{,}6/5{,}2 = 0{,}5$
0	1	$2{,}4/4{,}8 = 0{,}5$
1	1	$2{,}4/4{,}8 = 0{,}5$

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

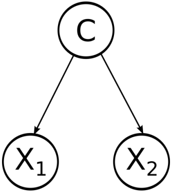


X_2	C	$p(X_2 \mid C)$
0	0	
1	0	
0	1	
1	1	

EM

Re-aprender el modelo

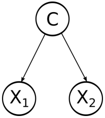
X_1	X_2	C = 0	C = 1
1	0	1.0	0.0
0	0	0.4	0.6
0	1	0.6	0.4
0	1	0.6	0.4
1	0	0.0	1.0
0	0	0.0	1.0
1	1	0.4	0.6
1	0	0.6	0.4
0	1	1.0	0.0
1	1	0.6	0.4



X_2	C	$p(X_2\|C)$
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

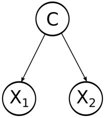
X_i	C	$p(X_1 \mid C)$	$p(X_2 \mid C)$
0	0	0.5	0.385
1	0	0.5	0.615
0	1	0.5	0.625
1	1	0.5	0.375

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

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

EM

Re-estimación de pesos



X_1	X_2	C	
		0	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

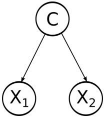
X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.5	0.385
1	0	0.5	0.615
0	1	0.5	0.625
1	1	0.5	0.375

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

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

EM

Re-estimación de pesos



X_1	X_2	C	
		0	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

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

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

EM

En la práctica...

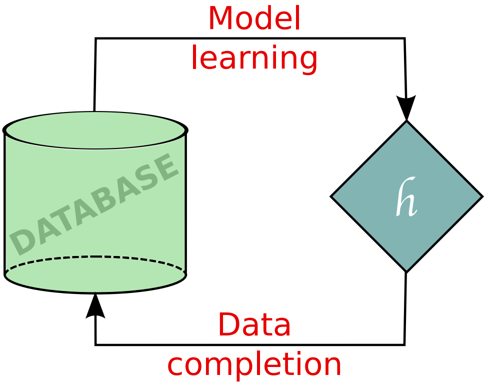
Estimación de parámetros: Laplace Smoothing

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

Cálculo de probabilidades: cálculo logarítmico

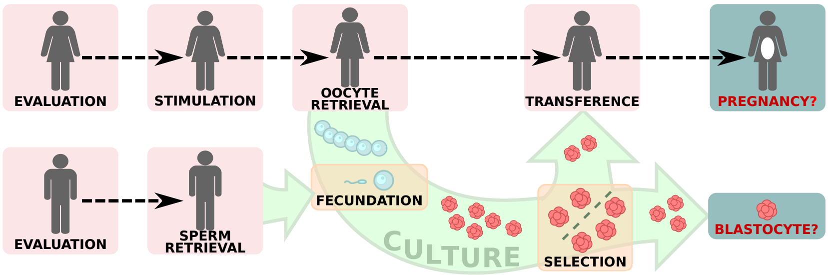
$$\hat{c} = \underset{c}{\operatorname{argmax}} \exp^{\left[\log p(c) + \sum_{i=1}^{2} \log p(x_i|c)\right]}$$

EM



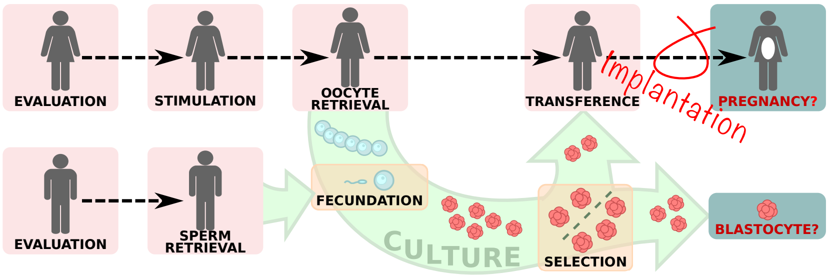
EM

Problema de la selección de embriones



EM

Problema de la selección de embriones

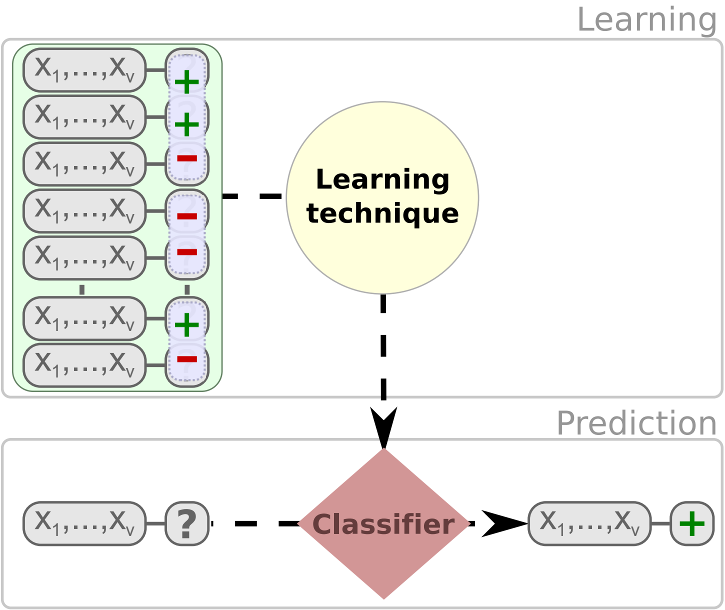


EM

Problema de la selección de embriones

EM

Problema de la selección de embriones



EM

Añadiendo información al proceso

X_1	X_2	C	lp
0	0	?	0.66
1	0	?	
0	1	?	
0	1	?	1.00
1	0	?	
1	0	?	0.33
1	0	?	
1	1	?	
0	1	?	0.50
1	1	?	

EM

Añadiendo información al proceso

X_1	X_2	C = 0	C = 1	lp
0	0	0.5	0.5	0.66
1	0	0.5	0.5	
0	1	0.5	0.5	
0	1	0.5	0.5	1.00
1	0	0.5	0.5	
1	0	0.5	0.5	0.33
1	0	0.5	0.5	
1	1	0.5	0.5	
0	1	0.5	0.5	0.50
1	1	0.5	0.5	

EM

Añadiendo información al proceso

X_1	X_2	C = 0	C = 1	lp
0	0	0.33	0.66	
1	0	0.33	0.66	0.66
0	1	0.33	0.66	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.66	0.33	
1	0	0.66	0.33	0.33
1	1	0.66	0.33	
0	1	0.50	0.50	0.50
1	1	0.50	0.50	

EM

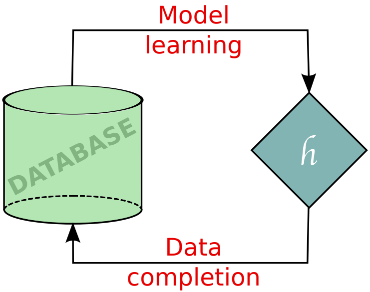
Añadiendo información al proceso

X_1	X_2	C = 0	C = 1	lp
0	0	0.33	0.66	0.66
1	0	0.33	0.66	
0	1	0.33	0.66	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.66	0.33	0.33
1	0	0.66	0.33	
1	1	0.66	0.33	
0	1	0.50	0.50	0.50
1	1	0.50	0.50	

EM

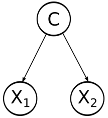
Añadiendo información al proceso

X_1	X_2	C = 0	C = 1	lp
0	0	0.33	0.66	0.66
1	0	0.33	0.66	
0	1	0.33	0.66	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.66	0.33	0.33
1	0	0.66	0.33	
1	1	0.66	0.33	
0	1	0.50	0.50	0.50
1	1	0.50	0.50	



EM informado

Aprendizaje del modelo



X_1	X_2	C = 0	C = 1
0	0	0.33	0.66
1	0	0.33	0.66
0	1	0.33	0.66
0	1	0.00	1.00
1	0	0.00	1.00
1	0	0.66	0.33
1	0	0.66	0.33
1	1	0.66	0.33
0	1	0.50	0.50
1	1	0.50	0.50

C	$p(C)$
0	0.4
1	0.6

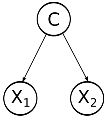
X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.29	0.50
1	0	0.71	0.50
0	1	0.47	0.50
1	1	0.53	0.50

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

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

EM informado

Aprendizaje del modelo



X_1	X_2	$C=0$	$C=1$
0	0	0.29	0.71
1	0	0.47	0.53
0	1	0.29	0.71
0	1	0.00	1.00
1	0	0.00	1.00
1	0	0.47	0.53
1	0	0.47	0.53
1	1	0.47	0.53
0	1	0.29	0.71
1	1	0.47	0.53

C	$p(C)$
0	0.4
1	0.6

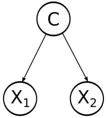
X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.29	0.50
1	0	0.71	0.50
0	1	0.47	0.50
1	1	0.53	0.50

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

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

EM informado

Aprendizaje del modelo



C	$p(C)$
0	0.4
1	0.6

X_i	C	$p(X_1\|C)$	$p(X_2\|C)$
0	0	0.29	0.50
1	0	0.71	0.50
0	1	0.47	0.50
1	1	0.53	0.50

X_1	X_2	C = 0	C = 1	lp
0	0	0.29	0.71	
1	0	0.47	0.53	0.66
0	1	0.29	0.71	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.47	0.53	
1	0	0.47	0.53	0.33
1	1	0.47	0.53	
0	1	0.29	0.71	0.50
1	1	0.47	0.53	

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

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

EM informado

Nuevo Paso-E

X_1	X_2	C = 0	C = 1	lp
0	0	0.29	0.71	
1	0	0.47	0.53	0.66
0	1	0.29	0.71	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.47	0.53	
1	0	0.47	0.53	0.33
1	1	0.47	0.53	
0	1	0.29	0.71	0.50
1	1	0.47	0.53	

Nuevo peso: suma ponderada de la probabilidad de cada asignación en las complexiones consistentes con las LPs.

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

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

EM informado

Nuevo Paso-E

X_1	X_2	C = 0	C = 1	lp
0	0	0.29	0.71	0.66
1	0	0.47	0.53	
0	1	0.29	0.71	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.47	0.53	0.33
1	0	0.47	0.53	
1	1	0.47	0.53	
0	1	0.29	0.71	0.50
1	1	0.47	0.53	

Nuevo peso: suma ponderada de la probabilidad de cada asignación en las complexiones consistentes con las LPs.

Complexiones

−	−	−	+	−	+	+	+
−	−	+	−	+	−	+	+
−	+	−	−	+	+	−	+

EM informado

Nuevo Paso-E

X_1	X_2	C = 0	C = 1	lp
0	0	0.29	0.71	
1	0	0.47	0.53	0.66
0	1	0.29	0.71	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.47	0.53	
1	0	0.47	0.53	0.33
1	1	0.47	0.53	
0	1	0.29	0.71	0.50
1	1	0.47	0.53	

Nuevo peso: suma ponderada de la probabilidad de cada asignación en las complexiones consistentes con las LPs.

Complexiones consistentes

				$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	
−	−	−	+	−	+	+	+
−	−	+	−	+	−	+	+
−	+	−	−	+	+	−	+
				$p(\mathbf{e}_1)$	$p(\mathbf{e}_2)$	$p(\mathbf{e}_3)$	

$$p(\mathbf{e}) = \prod_{j=1}^{N_i} p(C = e_j|\mathbf{x})$$

EM informado

Nuevo Paso-E

X_1	X_2	C = 0	C = 1	lp
0	0	E_{1-}	0.71	0.66
1	0	0.47	0.53	
0	1	0.29	0.71	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.47	0.53	0.33
1	0	0.47	0.53	
1	1	0.47	0.53	
0	1	0.29	0.71	0.50
1	1	0.47	0.53	

Nuevo peso: suma ponderada de la probabilidad de cada asignación en las complexiones consistentes con las LPs.

Complexiones consistentes

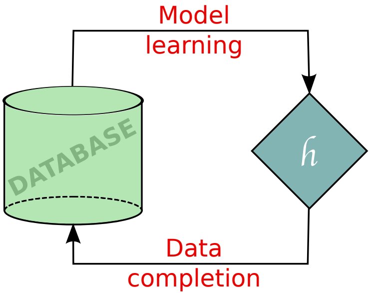
				$\mathbf{e}_1$	$\mathbf{e}_2$	$\mathbf{e}_3$	
−	−	−	+	−	+	+	+
−	−	+	−	+	−	+	+
−	+	−	−	+	+	−	+
				$p(\mathbf{e}_1)$	$p(\mathbf{e}_2)$	$p(\mathbf{e}_3)$	
				0,11	0,27	0,11	

$$E_{1-} = \frac{\sum_{t=1}^{S} p(\mathbf{e_t}) \cdot \mathbb{I}[e_{t1} = -]}{\sum_{t=1}^{S} p(\mathbf{e_t})}$$

EM informado

Nuevo Paso-E

X_1	X_2	C = 0	C = 1	lp
0	0	0.24	0.76	0.66
1	0	0.52	0.48	
0	1	0.24	0.76	
0	1	0.00	1.00	1.00
1	0	0.00	1.00	
1	0	0.66	0.33	0.33
1	0	0.66	0.33	
1	1	0.66	0.33	
0	1	0.32	0.68	0.50
1	1	0.68	0.32	



Aprendiendo con incertidumbre
Expectation-Maximization algorithm

Jerónimo Hernández González

Exploración y análisis de datos

Máster Universitario en Ingeniería
Computacional y Sistemas Inteligentes

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

MDE
Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School