Modelos Mixtos Gaussianos

Modelo $\longrightarrow \begin{cases} Z \sim \text{Cat}(\pi_1, \dots, \pi_K) \\ X \mid Z=j \sim f_j(x, \mu_j, \sigma_j) \end{cases}$

Vamos a resolver el ejemplo en el caso de gaussianas unidimensionales para fijar ideas.

La marginal de X

$$f_X(x,\theta) = \sum_{j=1}^{K} \pi_j f_j(x, \mu_j, \sigma_j) \quad , \quad f_j \sim N(\mu_j, \sigma_j) \ 1 \le j \le K$$

$$\theta = (\pi_1, \dots, \pi_K, \mu_1, \dots, \mu_K, \sigma_1, \dots, \sigma_K)$$

EM: $\quad X_1, \dots, X_n \quad$ i.i.d

$\qquad Z_1, \dots, Z_n \quad$ i.i.d

$$\text{loglik}(\theta) = \log\left(\prod_{i=1}^{n} f_{X_i}(x_i,\theta)\right) = \sum_{i=1}^{n} \log\left(f_{X_i}(x_i,\theta)\right) =$$

$$= \sum_{i=1}^{n} \log\left(\sum_{j=1}^{K} f_{X_i Z_i}(x_i, z_{ij}, \theta)\right) = (*)$$

Sea q una distribución cualquiera sobre Z_i

$$(*) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{K} q(z_{ij}) \frac{f_{X_i Z_i}(x_i, z_{ij}, \theta)}{q(z_{ij})}\right) \overset{\text{JENSEN}}{\geq}$$

$$\geq \sum_{i=1}^{n} \sum_{j=1}^{K} q(z_{ij}) \log\left(\frac{f_{X_i Z_i}(x_i, z_{ij}, \theta)}{q(z_{ij})}\right) = J(q,\theta)$$

Observamos que si tomo

$$\tilde{q}(z_{ij}) = P(Z_i = z_{ij} \mid X_i = x_i) \qquad 1 \le j \le K$$

Me queda

$$\tilde{q}(z_{ij}) = \frac{f_{X_i Z_i}(x_i, z_{ij}, \theta)}{f_{X_i}(x_i, \theta)} =$$

$$= \frac{f_{X_i | Z_i = z_{ij}}(x_i, \theta)\, P(Z_i = z_{ij})}{f_{X_i}(x_i, \theta)}$$

$$\tilde{q}(z_{ij}) = \frac{\pi_j\, f_j(x_i, \mu_j, \sigma_j)}{\sum_{l=1}^{K} \pi_l\, f_l(x_i, \mu_l, \sigma_l)} = w_{ij}(\theta)$$

Tomo parámetros iniciales

$$\theta^0 = (\pi_1^0, \dots, \pi_K^0, \mu_1^0, \dots, \mu_K^0, \sigma_1^0, \dots, \sigma_K^0) = (\pi^0, \mu^0, \sigma^0)$$

Aplico E-STEP: $\tilde{q}^0$ maximiza la $J(q, \theta^0)\ \forall q$

$$J(\tilde{q}^0, \theta^0) = \sum_{i=1}^{m} \sum_{j=1}^{K} \boxed{w_{ij}(\theta^0)} \log\left(f_{X_i}(x_i, \theta^0)\right)$$

Voy a dejar estos pesos fijos

Aplico M-STEP:

$J(\tilde{q}^0, \theta)$ es lo que debo maximizar

$$J(\tilde{q}^0, \theta) = \sum_{i=1}^{m} \sum_{j=1}^{K} w_{ij}(\theta^0) \log\left(\frac{\pi_j\, f_j(x_i, \mu_j, \sigma_j)}{w_{ij}(\theta^0)}\right) =$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{K} w_{ij}(\theta^0) \left[\log(\pi_j) + \log(f_j(x_i, \mu_j, \sigma_j)) - \log(w_{ij}(\theta^0))\right]$$

$$= \sum_{j=1}^{K} \sum_{i=1}^{m} w_{ij}(\theta^0) \log(\pi_j) + w_{ij}(\theta^0) \log(f_j(x_i, \mu_j, \sigma_j)) - w_{ij}(\theta^0) \log(w_{ij}(\theta^0))$$

Me queda una maximización con una restricción

$$\sum_{j=1}^{K} \pi_j = 1$$

Para resolver esto uso multiplicadores de Lagrange

$$L(\theta, \lambda) = J(\tilde{q}^0, \theta) - \lambda \left(\sum_{j=1}^{K} \pi_j - 1 \right)$$

Recordamos que los puntos críticos se encuentran donde $\nabla L(\theta, \lambda) = 0$

Calculamos las derivadas

$$1 \leq l \leq K \begin{cases} \dfrac{\partial}{\partial \mu_l} L(\theta,\lambda) = \displaystyle\sum_{i=1}^{m} w_{ij}(\theta^0) \frac{\partial}{\partial \mu_l} \log\left(f_l(x_i, \mu_l, \sigma_l)\right) & (1) \\ \dfrac{\partial}{\partial \sigma_l} L(\theta,\lambda) = \displaystyle\sum_{i=1}^{m} w_{ij}(\theta^0) \frac{\partial}{\partial \sigma_l} \log\left(f_l(x_i, \mu_l, \sigma_l)\right) & (2) \\ \dfrac{\partial}{\partial \pi_l} L(\theta,\lambda) = \displaystyle\sum_{i=1}^{m} w_{il}(\theta^0) \frac{1}{\pi_l} - \lambda & (3) \end{cases}$$

(3) $\dfrac{\partial}{\partial \pi_l} L(\theta, \lambda) = 0 \Rightarrow \pi_l = \displaystyle\sum_{i=1}^{m} \frac{w_{il}(\theta^0)}{\lambda} \qquad 1 \leq l \leq K$

Como $\displaystyle\sum_{l=1}^{K} \pi_l = 1 \Rightarrow \sum_{l=1}^{K} \sum_{i=1}^{m} \frac{w_{il}(\theta^0)}{\lambda} = 1$

$$\Rightarrow \lambda = \sum_{l=1}^{K} \sum_{i=1}^{m} w_{il}(\theta^0) = \sum_{i=1}^{m} \underbrace{\boxed{\sum_{l=1}^{K} w_{il}(\theta^0)}}_{=1} = m$$

Finalmente, $\boxed{\hat{\pi}_l = \dfrac{\sum_{i=1}^{m} w_{il}(\theta^0)}{m}}$

Para (1) y (2)

$$f_\ell(t, \mu_\ell, \sigma_\ell) = \frac{1}{\sqrt{2\pi\sigma_\ell^2}} e^{-\frac{1}{2}\frac{(t-\mu_\ell)^2}{\sigma_\ell^2}}$$

$$\log(f_\ell(t, \mu_\ell, \sigma_\ell)) = -\frac{1}{2}\log(2\pi\sigma_\ell^2) - \frac{1}{2}\frac{(t-\mu_\ell)^2}{\sigma_\ell^2}$$

Entonces,

$$\frac{\partial}{\partial \mu_\ell} J(\tilde{q}^0, \theta) = \sum_{i=1}^{n} w_{i\ell}(\theta^0) \frac{1}{\sigma^2}(x_i - \mu_\ell)$$

Igualando a cero y despejando

$$\boxed{\hat{\mu}_\ell = \sum_{i=1}^{n} \frac{w_{i\ell}(\theta^0)\, x_i}{w_{i\ell}(\theta^0)}}$$

$$\frac{\partial}{\partial \sigma_\ell} J(\tilde{q}^0, \theta) = \sum_{i=1}^{n} w_{i\ell}(\theta^0)\left[-\frac{1}{\sigma_\ell} + \frac{(x_i-\mu_\ell)^2}{\sigma_\ell^3}\right]$$

$$= \frac{1}{\sigma_\ell}\left[-\sum_{i=1}^{n} w_{i\ell}(\theta^0) + \frac{1}{\sigma_\ell^2}\sum_{i=1}^{n} w_{i\ell}(\theta^0)(x_i-\mu_\ell)^2\right]$$

Igualando a cero, despejando y reemplazando μ_ℓ por $\hat{\mu}_\ell$.

$$\boxed{\hat{\sigma}_\ell^2 = \sum_{i=1}^{n} \frac{w_{i\ell}(\theta^0)(x_i-\hat{\mu}_\ell)^2}{\sum_{i=1}^{n} w_{i\ell}(\theta^0)}}$$

Estas formulas se usan para iterar como vimos en la clase