

## 9. Simplificació de la barreja de Gaussians 1 [G]

Josep de Cid

Albert Ribes

Kerstin Winter

11 d'octubre de 2017

**Considereu el model de barreja de Gaussians:**

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x; \mu_k, \Sigma_k)$$

**Preneu el cas que totes les matrius de covariància són iguals i diagonals, és a dir,  $\Sigma_1 = \dots, \Sigma_K = \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$**

1. **Enraoneu en quin sentit representa una simplificació respecte al cas general (amb matrius de covariància generals), des dels punts de vista estadístic i geomètric.**

Significa que las dimensiones son independientes entre ellas.

Geomètricament esto significa que cada uno de los clusters generará instancias formando una probabilidad más circular, y menos ovalada.

2. **Expresseu la funció de densitat de probabilitat  $\mathcal{N}(x; \mu_k, \Sigma_k)$  que en resulta.**

En el caso genérico la función de densidad de la distribución gaussiana multivariada con  $D$  variables es:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{(\frac{D}{2})}} \cdot \frac{1}{(|\Sigma|)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

Pero se puede simplificar con las siguientes propiedades:

- $\det(\text{diag}(\alpha_1, \alpha_2, \dots, \alpha_n)) = \prod_{i=1}^n \alpha_i$
- $(\alpha_1, \alpha_2, \dots, \alpha_n)^T \cdot \text{diag}(\beta_1, \beta_2, \dots, \beta_n) = (\alpha_1\beta_1, \alpha_2\beta_2, \dots, \alpha_n\beta_n)^T$

Y entonces la fórmula anterior queda:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{(\frac{D}{2})}} \cdot \frac{1}{(\prod_{i=1}^n \sigma_i)} \exp\left(-\frac{1}{2} \sum_{i=1}^d \left(\frac{(x_i - \mu_i)^2}{\sigma_i^2}\right)\right)$$

3. **Construïu la funció de log-versemblança negativa.**

En el caso de que se disponga de  $M$  puntos de la distribución:

$$l = -\ln \prod_{j=1}^M p(x_j)$$

$$l = -\sum_{j=1}^M \ln(p(x_j))$$

$$l = -\sum_{j=1}^M \left[ \ln \left( \sum_{k=1}^K \pi_k \mathcal{N}(x_j; \mu_k, \Sigma_k) \right) \right]$$

$$l = -\sum_{j=1}^M \left[ \ln \left( \sum_{k=1}^K \left( \pi_k \frac{1}{(2\pi)^{(\frac{D}{2})}} \cdot \frac{1}{(\prod_{i=1}^n \sigma_i)} \exp \left( -\frac{1}{2} \sum_{i=1}^d \left( \frac{(x_{ji} - \mu_i)^2}{\sigma_i^2} \right) \right) \right) \right) \right]$$

**4. Deriveu les equacions de l'algorisme E-M que en resulta i escriviu l'algorisme de clustering complet.**

Hay que derivar respecto a  $\pi$ ,  $\mu$  y  $\Sigma$ .

La derivada respecto a cada una de las  $\pi_q$  es:

$$\frac{\partial l}{\partial \pi_q} = -\sum_{j=1}^M \left[ \frac{1}{\sum_{k=1}^K \pi_k \mathcal{N}(x_j; \mu_k, \Sigma_k)} \cdot \frac{\partial \left( \sum_{k=1}^K \pi_k \mathcal{N}(x_j; \mu_k, \Sigma_k) \right)}{\partial \pi_q} \right]$$

$$\frac{\partial l}{\partial \pi_q} = -\sum_{j=1}^M \left[ \frac{1}{\sum_{k=1}^K \pi_k \mathcal{N}(x_j; \mu_k, \Sigma_k)} \cdot \mathcal{N}(x_j, \mu_q, \Sigma_q) \right]$$

$$\frac{\partial l}{\partial \pi_q} = -\sum_{j=1}^M \left[ \frac{\mathcal{N}(x_j, \mu_q, \Sigma_q)}{\sum_{k=1}^K \pi_k \mathcal{N}(x_j; \mu_k, \Sigma_k)} \right]$$

$$\frac{\partial l}{\partial \pi_q} = -\sum_{j=1}^M \left[ \frac{p(x_j | \pi_k = \pi_q)}{p(x_j)} \right]$$

Usando el teorema de Bayes, que dice:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Podemos sustituirlo por:

$$\frac{\partial l}{\partial \pi_q} = -\sum_{j=1}^M \left[ \frac{\frac{p(\pi_k = \pi_q | x_j) p(x_j)}{p(\pi_k = \pi_q)}}{p(x_j)} \right]$$

$$\frac{\partial l}{\partial \pi_q} = - \sum_{j=1}^M \left[ \frac{p(\pi_k = \pi_q | x_j)}{p(\pi_k = \pi_q)} \right]$$

5. **Enraoneu sobre les implicacions (possibles avantatges/inconvenients) que representa la simplifcació respecte el cas general des del punt de vista del clustering.**