

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

Albert Ribes

10 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 covariança 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 covariança 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 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 ahora

3. **Construiu la funció de log-versemblança negativa.**
4. **Deriveu les equacions de l'algorisme E-M que en resulta i escriviu l'algorisme de clustering complet.**
5. **Enraoneu sobre les implicacions (possibles avantatges/inconvenients) que representa la simplificació respecte el cas general des del punt de vista del clustering.**