PSI3501 - Aprendizagem de Máquina e Processamento de Voz

Miguel Arjona Ramírez

O algoritmo EM geral para GMM

Log verossimilhança dos dados completos

$$L \equiv \ln p(X,Z|\theta)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \left[\ln \pi_k + \ln \underbrace{\mathcal{N}(x_n|\mu_k,\Sigma_k)}_{\mathcal{N}_k} \right]$$

$$L = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \left[\ln \pi_k + \ln \mathcal{N}_k \right]$$

— × —

Porque a verossimilhança conjunta é

$$\ln p(X,Z|\theta) = \sum_{n=1}^{N} \ln \prod_{k=1}^{K} \pi_k^{z_{nk}} \mathcal{N}^{z_{nk}}(x_n|\mu_k,\Sigma_k)$$

Como não dispomos efetivamente de $\{X, Z\}$, os dados completos, não podemos maximizar diretamente a log-verossimilhança conjunta $\ln p(X, Z \mid \theta)$.

Mas podemos maximizar sua esperança

$$Q(\theta, \theta^{ant}) = \sum_{Z} p(Z \mid X, \theta^{ant}) \ln p(X, Z \mid \theta)$$

em que θ^{ant} são os parâmetros atuais e θ são os novos parâmetros a calcular através da maximização

$$\theta^{novo} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta, \theta^{ant}).$$

[Na etapa M do algoritmo EM]

Já determinamos formalmente a densidade condicional a posteriori

$$p(x|z_k=1)=\mathcal{N}(x|\mu_k,\Sigma_k)$$

ou vetorialmente

$$p(x|z)=\prod_{k=1}^{K}\mathcal{N}^{z_k}(x|\mu_k,\Sigma_k)$$

e as densidades marginais

$$\begin{cases} p(z)=\prod_{k=1}^{K}\pi_k^{z_k} \\ p(x)=\sum_{j=1}^{K}\pi_j\mathcal{N}(x|\mu_j,\Sigma_j) \end{cases}$$

Como

$$p(z|x)\,p(x)=p(x|z)\,p(z)$$

que resulta no Teorema de Bayes com a inversão do condicionamento

$$p(z|x)=\frac{p(x|z).p(z)}{p(x)}$$

$$=\frac{\prod_{k=1}^{K}\pi_k^{z_k}\mathcal{N}^{z_k}(x|\mu_k,\Sigma_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x|\mu_j,\Sigma_j)}$$

Densidade a posteriori da variável latente

$$p(Z|X,\theta) = \prod_{n=1}^{N} \frac{\prod_{k=1}^{K} \left[\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)\right]^{z_{nk}}}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)}$$

que fatora em n, obtendo

$$p(z_{nk}|x_n,\theta_k) = \frac{\left[\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)\right]^{z_{nk}}}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)}$$

O valor esperado do indicador z_{nk} é

$$\gamma(z_{nk}) \equiv E[z_{nk}]$$

$$= \frac{\sum_{z_{nk}=0}^{1} z_{nk} \left[\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)\right]^{z_{nk}}}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)}$$

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n|\mu_k,\Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)}$$

O valor esperado da log verossimilhança do conjunto de dados completos é

$$\overline{L} = \mathbb{E}_{Z} \ln p(X, Z \mid \Theta)$$

$$= \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left[\ln \pi_k + \ln \mathcal{N}(\mu_k, \Sigma_k) \right]$$

$$\overline{L} = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left[\ln \pi_k + \ln \mathcal{N}_k \right]$$

EM geral para GMM

→ Vetores-média (Etapa M)

$$\mathcal{L} = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left[\ln \pi_k + \underbrace{\ln \mathcal{N}_k}_{\ln F_k + A_k} \right]$$

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{n=1}^{N} \gamma_{nk} \frac{\partial A_k}{\partial \mu_k}$$

$$\frac{\partial A_k}{\partial \mu_k} = \frac{\partial}{\partial \mu_k} \left[-\frac{1}{2} (x_n - \mu_k)^T \Sigma_k^{-1} (x_n - \mu_k) \right]$$

$$= \Sigma_k^{-1} (x_n - \mu_k)$$

$$\therefore \frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{n=1}^{N} \gamma_{nk} \Sigma_k^{-1} (x_n - \mu_k)$$

$$\frac{\partial \mathcal{L}}{\partial \mu_k} = \sum_{n=1}^{N} \gamma_{nk} \Sigma_{k\,\text{nova}}^{-1} \left(x_n - \mu_k^{\text{novo}} \right)$$

$$= \vec{0}$$

$$\Sigma_{k\,\text{nova}}^{-1} \sum_{n=1}^{N} \gamma_{nk} \left(x_n - \mu_k^{\text{novo}} \right) = \vec{0}$$

$$\sum_{n=1}^{N} \gamma_{nk} \left(x_n - \mu_k^{\text{novo}} \right) = \vec{0}$$

$$\mu_k^{\text{novo}} \sum_{n=1}^{N} \gamma(z_{nk}) = \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

$$\therefore \mu_k^{\text{novo}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

$\rightarrow$ Matrizes de covariância

$$\frac{\partial L}{\partial \Sigma_k} = \sum_{n=1}^{N} \gamma_{nk} \left[\frac{\partial \ln F_k}{\partial D_k} \cdot \frac{\partial D_k}{\partial \Sigma_k} + \frac{\partial A_k}{\partial \Sigma_k} \right]$$

$$\frac{\partial \ln F_k}{\partial D_k} = -\frac{1}{F_k D_k^2}, \quad D_k = (2\pi)^{D/2} \det^{1/2}(\Sigma_k)$$

$$\frac{\partial D_k}{\partial \Sigma_k} = (2\pi)^{D/2} \left(\frac{1}{2}\right) \det^{-1/2}(\Sigma_k) \cdot \frac{\partial}{\partial \Sigma_k} \det(\Sigma_k)$$

$$= (2\pi)^{D/2} \frac{1}{2} \det^{-\frac{1}{2}}(\Sigma_k) \cdot \det(\Sigma_k) \Sigma_k^{-1}$$

$$= \frac{1}{2} D_k \Sigma_k^{-1}$$

$$\frac{\partial A_k}{\partial \Sigma_k} = \frac{\partial}{\partial \Sigma_k}\left[-\frac{1}{2}(x_n-\mu_k)^T \Sigma_k^{-1}(x_n-\mu_k)\right]$$

$$= -\frac{1}{2}(x_n-\mu_k)(x_n-\mu_k)^T \frac{\partial \Sigma_k^{-1}}{\partial \Sigma_k}$$

$$= \frac{1}{2}(x_n-\mu_k)(x_n-\mu_k)^T\left(\Sigma_k^{-1}\right)^2$$

$\therefore$

$$\sum_{n=1}^{N} \gamma_{nk}\left[-\frac{1}{D_k}\frac{1}{2}D_k \Sigma_k^{-1\,nova} + \frac{1}{2}(x_n-\mu_k^{novo})(x_n-\mu_k^{novo})^T\left(\Sigma_k^{-1\,nova}\right)^2\right] = 0_{D\times D}$$

$$\left(\sum_{n=1}^{N}\gamma_{nk}\right)\Sigma_k^{-1\,nova} = \sum_{n=1}^{N}\gamma_{nk}(x_n-\mu_k^{novo})(x-\mu_k^{novo})^T\left(\Sigma_k^{-1\,nova}\right)^2$$

Multiplicando à direita por $\Sigma_k^{2\,nova}$ os dois membros da equação, obtemos

$$\Sigma_k^{nova} = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k^{novo})(x_n-\mu_k^{novo})^T$$

→ Pesos das componentes da mistura (Etapa M)

$$\mathcal{L} = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \left[\ln \pi_k + \ln F_k + A_k \right]$$

com a restrição

$$\sum_{k=1}^{K} \pi_k - 1 = 0$$

para compor o lagrangiano

$$J = \mathcal{L} + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

com condição de máximo

$$\frac{\partial J}{\partial \pi_k} = \sum_{n=1}^{N} \gamma_{nk} \frac{1}{\pi_k^{novo}} + \lambda = 0$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} \frac{\pi_k^{novo}}{\pi_k^{novo}} + \lambda \sum_{k=1}^{K} \pi_k^{novo} = 0$$

$$\lambda = -N$$

$$N_k \cdot \frac{1}{\pi_k^{novo}} - N = 0$$

$$\therefore \quad \pi_k^{novo} = \frac{N_k}{N}.$$