原型聚类 — 高斯混合聚类

Recall MLE (最大似然估计)

样本集 D，样本概率密度函数 $p(\vec{x}\mid\vec{\theta})$. ($\vec{\theta}$ 包含 $\vec{\mu}, \Sigma$，在此例中)

似然函数 $L(\vec{\theta}\mid D) = p(D\mid\vec{\theta}) = p(\vec{x}_1, \vec{x}_2, \cdots \vec{x}_m\mid\vec{\theta}) \overset{\text{i.i.d.}}{=\!=} \underline{\prod_{j=1}^{m} p(\vec{x}_i\mid\vec{\theta})}$

高斯混合分布 $\underline{p_M(\vec{x})} = \sum_{i=1}^{k} \alpha_i \, p(\vec{x}\mid\vec{\mu}_i, \Sigma_i) \quad \left(\sum_{i=1}^{k} \alpha_i = 1 \right)$

$\vec{\mu}_i, \Sigma_i$ 是第 i 个高斯分布的参数.

后验概率 $\gamma_{ji} = p_M(z_j = i\mid\vec{x}_j) = \dfrac{P(z_j=i)\cdot p_M(\vec{x}_j\mid z_j=i)}{p_M(\vec{x}_j)}$

$$= \frac{\alpha_i \cdot p(\vec{x}_j\mid\vec{\mu}_i, \Sigma_i)}{\sum_{l=1}^{k} \alpha_l \cdot p(\vec{x}_j\mid\vec{\mu}_l, \Sigma_l)}$$

似然函数 $LL(D) = \prod_{j=1}^{m} p_M(\vec{x}_j) = \prod_{j=1}^{m}\sum_{i=1}^{k} \alpha_i \cdot p(\vec{x}_j\mid\vec{\mu}_i, \Sigma_i)$

右两边取 ln. $\Rightarrow$ $LL(D) = \sum_{j=1}^{m} \ln\left(\sum_{i=1}^{k} \underline{\alpha_i} \cdot p(\vec{x}_j\mid\underline{\vec{\mu}_i}, \underline{\Sigma_i}) \right)$ 参数 (变量)

最大化 $LL(D)$. 求 $\boxed{\dfrac{\partial LL(D)}{\partial \vec{\mu}_i}}$

首先 $p(\vec{x}_j\mid\vec{\mu}_i, \Sigma_i) = \dfrac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x}_j - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i)\right\}$

注意: <u>当前只有 $\vec{\mu}_i$ 为变量</u>. $\dfrac{\partial LL(D)}{\partial \vec{\mu}_i} = \dfrac{\partial\left\{\sum_{j=1}^{k} \ln\left(\sum_{i=1}^{k}\alpha_i \cdot p(\vec{x}_j\mid\vec{\mu}_i, \Sigma_i)\right)\right\}}{\partial \vec{\mu}_i}$

$$\therefore \frac{\partial LL(D)}{\partial \vec{\mu}_i} = \sum_{j=1}^{K} \frac{\partial \ln\left(\sum_{i=1}^{K} \alpha_i \cdot p(\vec{x}_j \mid \vec{\mu}_i, \Sigma_i)\right)}{\partial \vec{\mu}_i} = \sum_{j=1}^{K} \frac{1}{\sum_{i=1}^{K} \alpha_i\, p(\vec{x}_j \mid \vec{\mu}_i, \Sigma_i)} \cdot \underbrace{\frac{\partial(\alpha_i\, p(\vec{x}_j \mid \vec{\mu}_i, \Sigma_i))}{\partial \vec{\mu}_i}}_{\beta}$$

$$\beta = \frac{\partial(\alpha_i \cdot p(\vec{x}_j \mid \vec{\mu}_i\, \Sigma_i))}{\partial \vec{\mu}_i} = \alpha_i \cdot \frac{\partial\, p(\vec{x}_j \mid \vec{\mu}_i\, \Sigma_i)}{\partial \vec{\mu}_i} = \alpha_i \cdot \underbrace{c}_{\frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}}}} \cdot \frac{\partial \exp\left\{-\frac{1}{2}(\vec{x}_j - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i)\right\}}{\partial \vec{\mu}_i}$$

$$= \alpha_i \cdot c \cdot \exp\left\{-\frac{1}{2}(\vec{x}_j - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i)\right\} \cdot \frac{\partial \left\{-\frac{1}{2}(\vec{x}_j - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i)\right\}}{\partial \vec{\mu}_i}$$

$$= \alpha_i \cdot p(\vec{x}_j \mid \vec{\mu}_i\, \Sigma_i) \cdot \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i)$$

$$\therefore \frac{\partial LL(D)}{\partial \vec{\mu}_i} = \sum_{j=1}^{K} \underbrace{\frac{\alpha_i \cdot p(\vec{x}_j \mid \vec{\mu}_i, \Sigma_i)}{\sum_{i=1}^{K} \alpha_i \cdot p(\vec{x}_j \mid \vec{\mu}_i, \Sigma_i)}}_{\gamma_{ji}} \cdot \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i) = 0$$

$$\therefore \sum_{j=1}^{K} \gamma_{ji}\, \Sigma_i^{-1} (\vec{x}_j - \vec{\mu}_i) = 0$$

$$\sum_{j=1}^{K} \gamma_{ji}\, \Sigma_i^{-1} \vec{\mu}_i = \sum_{j=1}^{K} \gamma_{ji}\, \Sigma_i^{-1} \vec{x}_j$$ 两边乘以 Σ_i^{-1}

$$\vec{\mu}_i = \frac{\sum_{j=1}^{K} \gamma_{ji}\, \vec{x}_j}{\sum_{j=1}^{K} \gamma_{ji}}$$

$$\therefore \Sigma_i = \frac{\sum_{j=1}^{K} \gamma_{ji} (\vec{x}_j - \vec{\mu}_i)(\vec{x}_j - \vec{\mu}_i)^T}{\sum_{j=1}^{K} \gamma_{ji}}$$

P36.

$$\sum_{j=1}^{m} \frac{p(\vec{x_j} \mid \vec{\mu_i}, \Sigma_i)}{\sum_{l=1}^{k} \alpha_l \cdot p(\vec{x_j} \mid \vec{\mu_i}, \Sigma_i)} + \lambda = 0$$

乘以 α_i

$$\because \gamma_{ji} = \frac{\alpha_i \cdot p(\vec{x_j} \mid \vec{\mu_i}, \Sigma_i)}{\sum_{l=1}^{k} \alpha_l \cdot p(\vec{x_j} \mid \vec{\mu_i}, \Sigma_i)}$$

相同.

$$\sum_{j=1}^{m} \gamma_{ji} + \alpha_i \lambda = 0$$

$$\therefore \alpha_i = -\frac{1}{\lambda} \sum_{j=1}^{m} \gamma_{ji}$$

$$1 = \sum_{i=1}^{k} \alpha_i = -\frac{1}{\lambda} \sum_{i=1}^{k} \sum_{j=1}^{m} \gamma_{ji} = -\frac{1}{\lambda} \underbrace{\sum_{j=1}^{m} \underbrace{\sum_{i=1}^{k} \gamma_{ji}}_{=1}}_{=m}$$

$$\therefore \lambda = -m$$

因此. $\alpha_i = \frac{1}{m} \sum_{j=1}^{m} \gamma_{ji}$