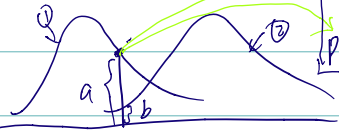
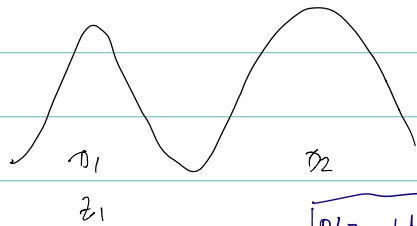


# 用EM解混合高斯



$$P(z_i=1|x_i, \theta) = \frac{a}{a+b}$$

$$P(z_i=2|x_i, \theta) = \frac{b}{a+b}$$

E-step

$$P(X, z|\theta) = \prod_{i=1}^n P(z_i, x_i|\theta) = \prod_{i=1}^n P(x_i|z_i, \theta) P(z_i|\theta) = \prod_{i=1}^n d_{z_i} N(\mu_{z_i}, \Sigma_{z_i})$$

$$\theta^{(g+1)} = \arg \max_{\theta} \int \log P(X, z|\theta) P(z|X, \theta^{(g)}) dz$$

$$\prod_{i=1}^n P(z_i|x_i, \theta^{(g)})$$

$$P(z_i|x_i, \theta^{(g)}) = \frac{P(z_i|x_i) P(x_i)}{\sum_{z_i=1}^k P(z_i|x_i) P(x_i)} = \frac{N(x_i|\mu_{z_i}, \Sigma_{z_i}) d_{z_i}}{\sum_{z_i=1}^k N(x_i|\mu_{z_i}, \Sigma_{z_i}) d_{z_i}}$$

$$\frac{k}{z_i=1} \dots \frac{k}{z_N=1} \left( \underbrace{\sum_{i=1}^N [\log d_{z_i} + \log N(x_i|\mu_{z_i}, \Sigma_{z_i})]}_{f(z_i)} \cdot \underbrace{\prod_{i=1}^N P(z_i|x_i, \theta^{(g)})}_{P(z_1, \dots, z_N)} \right)$$

$$= \sum_{z_i=1}^k \dots \sum_{z_N=1}^k (f_1(z_1) + f_2(z_2) + \dots + f_N(z_N)) \cdot P(z_1, \dots, z_N)$$

$$= \sum_{z_i=1}^k \dots \sum_{z_N=1}^k f_1(z_1) P(z_1, \dots, z_N) + \dots$$

$$= \sum_{z_i=1}^k f_1(z_i) \sum_{z_2=1}^k \dots \sum_{z_N=1}^k P(z_1, \dots, z_N) + \dots$$

$$= \sum_{z_i=1}^k f_1(z_i) P(z_i) + \dots$$

$$= \sum_{i=1}^N \sum_{z_i=1}^k f_i(z_i) P(z_i)$$

$$= \sum_{i=1}^N \sum_{z_i=1}^k (\log d_{z_i} + \log N(x_i|\mu_{z_i}, \Sigma_{z_i})) \cdot P(z_i|x_i, \theta^{(g)})$$

M-step

$$\sum_{i=1}^N \sum_{z_i=1}^k \log d_{z_i} \cdot P(z_i|x_i, \theta^{(g)}) = M$$

$$\frac{\partial M}{\partial \mu_1 \dots \partial \mu_N} = 0$$

相互迭代求解高斯

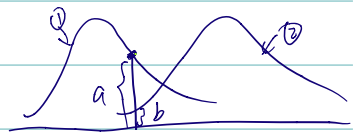
$$\text{最大化 } \mu, \Sigma \text{ 使 } \theta^{(g)} \rightarrow \theta^{(g+1)}$$

$$d_{z_i}^{(g+1)} = \frac{1}{N} \sum_{i=1}^N P(z_i|x_i, \theta^{(g)})$$

$$\mu_c^{(g+1)} = \frac{\sum_{i=1}^N x_i P(z_i|x_i, \theta^{(g)})}{\sum_{i=1}^N P(z_i|x_i, \theta^{(g)})}$$

某一个高斯分量的期望有  $x_i$

$$\Sigma_c^{(g+1)} = \frac{\sum_{i=1}^N [x_i - \mu_c][x_i - \mu_c]^T P(z_i|x_i, \theta^{(g)})}{\sum_{i=1}^N P(z_i|x_i, \theta^{(g)})}$$



$$\textcircled{1} \frac{\partial}{\partial x} x^T A x = \text{tr}(A \frac{\partial}{\partial x} x x^T)$$

$$\textcircled{2} \frac{\partial \text{tr}(x^T B x)}{\partial x} = 2x^T - \text{diag}(x^T)$$

③  $x$  对称

$$\frac{\partial \text{tr}(x^T B x)}{\partial x} = B + B^T - \text{diag}(B)$$

However, in general

$$\frac{\partial \text{tr}(x^T B x)}{\partial x} = B^T$$

# M-step 推导

- 主要是对 E-step 构造的式子进行参数最大化,

$$\frac{\partial \sum_{i=1}^k \sum_{l=1}^n \log \alpha_l P(z_i | x_i, \theta^{(q)})}{\partial \alpha_1 \dots \partial \alpha_k} = [0 \dots 0]$$

$$\text{subject: } \sum_{l=1}^k \alpha_l = 1 \text{ 及 } z_i = 1$$

拉格朗日乘子法

$\alpha_l$

$$J(\alpha_1, \dots, \alpha_k, \lambda) = \sum_{i=1}^k \log \alpha_l \sum_{l=1}^n P(l | x_i, \theta^{(q)}) - \lambda (\sum_{l=1}^k \alpha_l - 1)$$

$$\frac{\partial J}{\partial \alpha_l} = \frac{1}{\alpha_l} \sum_{l=1}^n P(l | x_i, \theta^{(q)}) - \lambda = 0$$

$$\alpha_l = \frac{1}{\lambda} \left( \sum_{l=1}^n P(l | x_i, \theta^{(q)}) \right)$$

$$\sum_{l=1}^k \alpha_l = \frac{1}{\lambda} \sum_{l=1}^k \sum_{l=1}^n P(l | x_i, \theta^{(q)})$$

$$\therefore \sum_{l=1}^k P(l | x_i, \theta^{(q)}) = 1 \quad \text{一个样本属于所有高斯的情况加和。}$$

$$\therefore \sum_{l=1}^k \alpha_l = \frac{\sum_{l=1}^n 1}{\lambda} = 1$$

$$\therefore \lambda = n$$

$$\therefore \alpha_l = \frac{\sum_{l=1}^n P(l | x_i, \theta^{(q)})}{n}$$

$$\textcircled{1} \sum_{i=1}^n x_i^T \Delta x_i = \text{tr}(\Delta \sum_{i=1}^n x_i x_i^T) \quad \mu_L$$

$$\textcircled{2} \frac{\partial \ln |X|}{\partial X} = 2X^T - \text{diag}(X^{-1})$$

③  $X$  对称

$$\frac{\partial \text{tr}(XB)}{\partial X} = B + B^T - \text{diag}(B)$$

However, in general

$$\frac{\partial \text{tr}(XB)}{\partial X} = B^T$$

我们可对 Q 项加推导的第二项分析。

$$S = \sum_{i=1}^k \sum_{l=1}^n (\log N(x_i | \mu_l, \Sigma_l)) P(z_i | x_i, \theta^{(q)})$$

$$= \sum_{i=1}^k \sum_{l=1}^n (\log N(x_i | \mu_l, \Sigma_l)) P(l | x_i, \theta^{(q)})$$

$$= \sum_{l=1}^n -\frac{1}{2} \log |\Sigma_l| - \frac{1}{2} (x_i - \mu_l)^T \Sigma_l^{-1} (x_i - \mu_l) P(l | x_i, \theta^{(q)})$$

$$\text{Fact1: } = -\text{Tr} \left( \frac{\Sigma_l^{-1}}{2} \sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)}) \right) + \text{const.}$$

$$\text{Fact2: } \frac{\partial S}{\partial \mu_l} = \frac{\sum_{i=1}^n \Sigma_l^{-1} (x_i - \mu_l) P(l | x_i, \theta^{(q)})}{2} = 0$$

$$\mu_l = \frac{\sum_{i=1}^n x_i P(l | x_i, \theta^{(q)})}{\sum_{i=1}^n P(l | x_i, \theta^{(q)})}$$

$\Sigma_L$  我们可对  $\Sigma_l$  求偏导, 再利用 Fact2.

$$\frac{\partial S}{\partial \Sigma_l} = \frac{\sum_{i=1}^n \Sigma_l^{-1} P(l | x_i, \theta^{(q)}) - \Sigma_l^{-1} \text{diag}(\Sigma_l) P(l | x_i, \theta^{(q)})}{2} - \frac{\sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)}) - \text{diag}(\sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)}))}{2}$$

$$= 2 \left( \sum_{i=1}^n \Sigma_l^{-1} P(l | x_i, \theta^{(q)}) - \Sigma_l^{-1} \text{diag}(\Sigma_l) P(l | x_i, \theta^{(q)}) \right) - \sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)}) - \sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)})$$

$$= \sum_{i=1}^n \Sigma_l^{-1} P(l | x_i, \theta^{(q)}) - \sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)}) = 0$$

$$\Sigma_l = \frac{\sum_{i=1}^n (x_i - \mu_l)(x_i - \mu_l)^T P(l | x_i, \theta^{(q)})}{\sum_{i=1}^n P(l | x_i, \theta^{(q)})}$$



