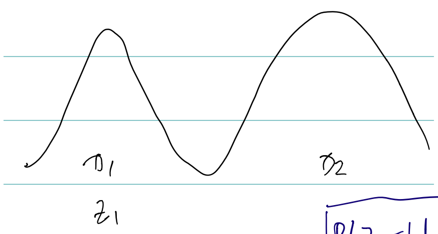
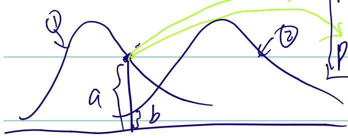
# 用EM 解 混合高斯




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

$$\theta^{(g+1)}=\arg\max_{\theta}\int_z \log P(X,z|\theta)\,\underbrace{P(z|X,\theta^{(g)})}_{\prod_{i=1}^{N}P(z_i|x_i,\theta^{(g)})}\,dz$$




$$\boxed{\begin{array}{l}P(z_i=1|x_i,\theta)=\frac{a}{a+b}\\ P(z_i=2|x_i,\theta)=\frac{b}{a+b}\end{array}}\rightarrow P(z_i|x_i,\theta^{(g)})=\frac{P(x_i|z_i)P(z_i)}{\sum_{z_i=1}^{k}P(x_i|z_i)P(z_i)}=\frac{N(x_i|\mu_{z_i},\Sigma_{z_i})\alpha_{z_i}}{\sum_{l=1}^{k}N(x_i|\mu_l,\Sigma_l)\alpha_l}$$

E-step

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

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

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

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

$$=\sum_{z_1=1}^{k}f_1(z_1)P(z_1)+\cdots$$

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

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

M-step

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

$$\frac{\partial M}{\partial\alpha_1\cdots\partial\alpha_N}=0$$ 相当于优化每个高斯.

① $\sum x_i^T A x_i=tr(A\sum x_i x_i^T)$

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

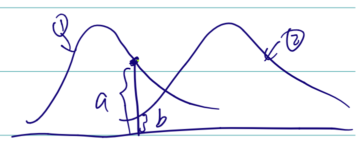
③ 当 X 对称

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

However, in general

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

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




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

$$\mu_l^{(g+1)}=\frac{\sum_{i=1}^{N}x_i P(l|x_i,\theta^{(g)})}{\sum_{i=1}^{N}P(l|x_i,\theta^{(g)})}$$ 某一个高斯下的所有 $x_i$

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

# M-step 推导

- 主要是对 E-step 构造的式子进行参数最大化,

$$\frac{\partial \sum_{l=1}^{k}\sum_{i=1}^{n} \log \alpha_l \, p(z_i \mid x_i, \theta^{(g)})}{\partial \alpha_1 \cdots \partial \alpha_k} = [0 \cdots 0]$$

subject: $\sum_{l=1}^{k} \alpha_l = 1$ 定义 $z_i = l$.

拉格朗日乘子法

$$L(\alpha_1, \ldots, \alpha_k, \lambda) = \sum_{l=1}^{k} \log \alpha_l \sum_{i=1}^{n} p(l \mid x_i, \theta^{(g)}) - \lambda\left(\sum_{l=1}^{k} \alpha_l - 1\right).$$

$\alpha_l$:

$$\frac{\partial L}{\partial \alpha_l} = \frac{1}{\alpha_l} \sum_{i=1}^{k} p(l \mid x_i, \theta^{(g)}) - \lambda = 0$$

$$\alpha_l = \frac{1}{\lambda}\left(\sum_{i=1}^{n} p(l \mid x_i, \theta^{(g)})\right).$$

$$\sum_{l=1}^{k} \alpha_l = \sum_{l=1}^{k} \frac{\sum_{i=1}^{n} p(l \mid x_i, \theta^{(g)})}{\lambda}$$

$\because \sum_{l=1}^{k} p(l \mid x_i, \theta^{(g)}) = 1$ 一个样子属于所有高斯的情况加和.

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

$\therefore \lambda = n$

$$\therefore \alpha_l = \frac{\sum_{i=1}^{n} p(l \mid x_i, \theta^{(g)})}{n}.$$

① $\sum x_i^T A x_i = tr(A \sum x_i x_i^T)$

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

③ 当 X 对称
$\frac{\partial tr(XB)}{\partial X} = B + B^T - diag(B)$

However, in general
$\frac{\partial tr(XB)}{\partial X} = B^T$

$\mu_l$: 我们对 Q 项加和里面的第二项分析.

$$S = \sum_{z_i=1}^{k} \sum_{i=1}^{n} \left(\log N(x_i \mid \mu_{z_i}, \Sigma_{z_i})\right) p(z_i \mid x_i, \theta^{(g)}).$$

$$= \sum_{l=1}^{k} \sum_{i=1}^{n} \left(\log N(x_i \mid \mu_l, \Sigma_l)\right) p(l \mid x_i, \theta^{(g)}).$$

$$= \sum_{i=1}^{n} -\frac{1}{2}\log(|\Sigma_l|) - \frac{1}{2}(x_i - \mu_l)^T \Sigma^{-1} (x_i - \mu_l)\, p(l \mid x_i, \theta^{(g)}).$$

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

Fact2. $:= \frac{\partial S}{\partial \mu_l} = \frac{2\Sigma_l^{-1} - diag(\Sigma_l^{-1})}{2} \sum_{i=1}^{n} 2(x_i - \mu_l)\, p(l \mid x_i, \theta^{(g)}) = 0$

$$\mu_l = \frac{\sum_{i=1}^{n} x_i \, p(l \mid x_i, \theta^{(g)})}{\sum_{i=1}^{n} p(l \mid x_i, \theta^{(g)})}$$

$\Sigma_l$: 我们对 $\Sigma_l^{-1}$ 求偏导, 再利用 Fact 2.

$$\frac{\partial S}{\partial \Sigma_l^{-1}} = \frac{2\sum_{i=1}^{n} \Sigma_l\, p(l \mid x_i, \theta^{(g)}) - \Sigma_i diag(\Sigma_l) p(l \mid x_i, \theta^{(g)})}{2} - \frac{2\Sigma_i (x_i - \mu_l)(x_i - \mu_l)^T p(l \mid x_i, \theta^{(g)}) - diag(\Sigma_i (x_i - \mu_l)(x_i - \mu_l)^T) p(l \mid x_i, \theta^{(g)})}{2}$$

$$= 2\left(\sum_{i=1}^{n} \Sigma_l\, p(l \mid x_i, \theta^{(g)}) - \sum_{i=1}^{n} (x_i - \mu_l)(x_i - \mu_l)^T\right) p(l \mid x_i, \theta^{(g)}) - \sum_{i=1}^{n} diag(\Sigma_l\, p(l \mid x_i, \theta^{(g)})) - \sum_{i=1}^{n} (x_i - \mu_l)(x_i - \mu_l)^T p.$$

$$= \sum_{i=1}^{n} \Sigma_l\, p(l \mid x_i, \theta^{(g)}) - \sum_{i=1}^{n} (x_i - \mu_l)(x_i - \mu_l)^T\, p(l \mid x_i, \theta^{(g)}) = 0$$

$$\Sigma_l = \frac{\sum_{i=1}^{n} (x_i - \mu_l)(x_i - \mu_l)^T p(l \mid x_i, \theta^{(g)})}{\sum_{i=1}^{n} p(l \mid x_i, \theta^{(g)})}$$