EM 算法

对于m个样本观察数据$x = (x^{(1)}, x^{(2)}, ... x^{(m)})$中，找出样本的模型参数$\theta$，极大化模型分布的对数似然函数如下，假设数据中有隐含变量$z = (z^{(1)}, z^{(2)}, ... z^{(m)})$

$$L(\theta) = \sum_{i=1}^{m} logP(x^{(i)}|\theta)$$

加入隐含变量公式变为如下，注意到下式中$Q_i(z(i))$是一个分布，因此$\sum Q_i(z(i))logP(x(i), z(i)|\theta)$可以理解为$logP(x(i), z(i)|\theta)$基于条件概率分布$Q_i(z(i))$的期望。

$$Q_i(z^{(i)}) = P(z^{(i)}|x^{(i)}, \theta)$$

$$L(\theta) = \sum_{i=1}^{m} log \sum_{z^{(i)}} Q_i(z^{(i)}) P(x^{(i)}, z^{(i)}|\theta) \quad s.t. \sum_{z} Q_i(z^{(i)}) = 1 \qquad (1)$$

根据Jensen不等式,(1)式变为(2)

$$E[f(g(X))] \geq f(E[g(X)])$$

$$L(\theta) = \sum_{i=1}^{m} log \sum_{z^{(i)}} Q_i(z^{(i)}) P(x^{(i)}, z^{(i)}|\theta) \geq \sum_{i=1}^{m} \sum_{z^{(i)}} Q_i(z^{(i)}) logP(x^{(i)}, z^{(i)}|\theta) \quad s.t. \sum_{z} Q_i(z^{(i)}) = 1 \qquad (2)$$

EM 直观

EM

- E-step:

$$Q(\theta \mid \theta^{(t)}) = \int \log L(\theta \mid Y)\, p(Z \mid X, \theta^{(t)})$$

$$= E[\log L(\theta \mid Y) \mid X, \theta^{(t)}]$$

$$= E[\log p(X, Z \mid \theta) \mid X, \theta^{(t)}]$$

M-step:

$$\theta^{(t+1)} = \underset{\theta}{\arg\max}\, Q(\theta \mid \theta^{(t)}).$$

停止条件:

$$|Q^{(t-1)} - Q^{(t)}|$$ 足够小.

逐渐逼近