EM 算法

如果概率模型的变量都是观测变量，那么给定数据，可以直接使用极大似然估计法。但当模型含有隐含变量时，不能简单的进行极大似然估计，需要EM算法。因此EM算法是对含有隐含变量的概率模型参数进行极大似然估计。

为了说明EM算法，首先介绍Jession不等式。

Jession inequality

如果 f 为凸函数$f''(x) \geq 0$,那么$f(E(x)) \leq E(f(x))$

如果 f 为凹函数$f''(x) \leq 0$,那么$f(E(x)) \geq E(f(x))$

$f(E(x)) = E(f(x))$ 的充要条件是随机变量x在概率1下取相同值。

下面来推导EM算法：

模型：$P(x, z, \theta)$

仅观测到 x

目标：最大化极大似然函数 $l(\theta) = \sum_{i=1}^{m} log(P(x^i, \theta)) = \sum_{i=1}^{m} log(\sum_{z^i} P(x^i, z^i \theta))$

上述问题的难点是log中含有加法。

基本精神 ：初始化参数θ，建立一个对数似然函数的比较紧密的下界，猜测参数之后，找到这个函数取最大值的θ，并重复上述过程，直到收敛到函数的一个局部最优值（不保证达到全局最优解）。

$l(\theta) = \sum_i log(P(x^i, \theta)) = \sum_{i=1} log(\sum_{z^i} P(x^i, z^i, \theta))$

$= \sum_{i=1} log(\sum_{z^i} Q(z^i) \frac{P(x^i, z^i, \theta))}{Q(z^i)}$

其中$Q_i(z^i)$为z^i的一个概率分布

$= \sum_i log E_{z^i \in Q} [\frac{P(x^i, z^i, \theta))}{Q(z^i)}]$

根据Jession不等式，有

$\geq \sum_i E_{z^i \in Q} [log(\frac{P(x^i, z^i, \theta))}{Q(z^i)})]$

$= \sum_i \sum_{z^i} Q(z^i) log(\frac{P(x^i, z^i, \theta))}{Q(z^i)})$

当$\frac{P(x^i, z^i, \theta))}{Q(z^i)}$以概率1取恒定值时，等号成立，因此，获得原始函数的一个紧密的下界。选取合适的$Q(z^i)$使得等式成立。

根据$\frac{P(x^i, z^i, \theta))}{Q_i(z^i)} = constant$,可以得到：$Q(z^i)$正比于$P(x^i, z^i, \theta)$

$\sum_{z^i} Q(z^i) = 1$

根据上边两式可以得到：$Q(z^i) = \frac{P(x^i, z^i, \theta)}{\sum_{z^i} P(x^i, z^i, \theta)} = \frac{P(x^i, z^i, \theta)}{P(x^i, \theta)} = P(z^i | x^i, \theta)$

因此EM算法总结如下：

E-step：

计算 $Q_i(z^i) = P(z^i|x^i, \theta)$

M-step:

最大化紧却下界函数：

$$\theta = argmax_\theta \sum_i \sum_{z^i} Q(z^i) log(\frac{P(x^i, z^i, \theta))}{Q(z^i)})$$