

# EM算法

$Y$ 表示能观测到的随机变量， $Z$ 表示隐变量。 $Y$ 和 $Z$ 连在一起的称为完全数据（complete-data），仅仅只有 $Y$ 称为不完全数据（incomplete-data）。 $P(Y|\Theta)$ 称为不完全数据的概率分布， $P(Y, Z|\Theta)$ 称为完全数据的概率分布。需要估计的模型参数是 $\Theta$ ，因此 $P(Y|\Theta)$ 称为不完全数据的似然函数， $P(Y, Z|\Theta)$ 称为完全数据的似然函数。

## EM算法的导出

$$\begin{aligned}\text{对数似然函数 } L(\Theta) &= \log P(Y|\Theta) \\ &= \log \sum_z (P(Y, Z|\Theta)) \quad (\text{应用边缘分布求和}) \\ &= \log \sum_z (P(Z|\Theta)P(Y|Z, \Theta)) \quad (\text{应用条件概率；此时为全概率公式})\end{aligned}$$

因为原始的问题包含一个隐变量 $Z$ ，所以这个对数似然函数公式中包含对 $Z$ 求和的计算。因为 $\log$ 函数中存在加法运算，因此想要通过求导来直接求解极大似然会很困难。EM算法不是采用直接求解，而是迭代的方法一步一步极大化 $L(\Theta)$ ，近似求解参数 $\Theta$ 的。记第 $i$ 次迭代之后的参数为 $\Theta^{(i)}$ ，我们计算 $L(\Theta)$ 与 $L(\Theta^{(i)})$ 的差值：

$$\begin{aligned}L(\Theta) - L(\Theta^{(i)}) &= \log P(Y|\Theta) - \log P(Y|\Theta^{(i)}) \\ &= \log \sum_z (P(Z|\Theta)P(Y|Z, \Theta)) - \log P(Y|\Theta^{(i)}) \\ &= \log \sum_z \left\{ P(Z|Y, \Theta^{(i)}) \frac{P(Z|\Theta)P(Y|Z, \Theta)}{P(Z|Y, \Theta^{(i)})} \right\} - \log P(Y|\Theta^{(i)})\end{aligned}$$

上式中红色的部分是我们新补充的项，构造了一个新的表达式。因为 $\log(x)$ 是凹函数，且 $\sum_z P(Z|Y, \Theta^{(i)}) = 1$ 。这样我们构造出的表达式就可以利用Jensen不等式：

$$\begin{aligned}L(\Theta) - L(\Theta^{(i)}) &= \log \sum_z \left\{ P(Z|Y, \Theta^{(i)}) \frac{P(Z|\Theta)P(Y|Z, \Theta)}{P(Z|Y, \Theta^{(i)})} \right\} - \log P(Y|\Theta^{(i)}) \\ &\geq \sum_z \left\{ P(Z|Y, \Theta^{(i)}) \log \frac{P(Z|\Theta)P(Y|Z, \Theta)}{P(Z|Y, \Theta^{(i)})} \right\} - \log P(Y|\Theta^{(i)}) \\ &= \sum_z \left\{ P(Z|Y, \Theta^{(i)}) \log \frac{P(Z|\Theta)P(Y|Z, \Theta)}{P(Z|Y, \Theta^{(i)}) \log P(Y|\Theta^{(i)})} \right\} \dots eq. (1)\end{aligned}$$

$$\begin{aligned}\text{令 } B(\Theta, \Theta^{(i)}) &= L(\Theta^{(i)}) + eq. (1) \\ &= L(\Theta^{(i)}) + \sum_z \left\{ P(Z|Y, \Theta^{(i)}) \log \frac{P(Z|\Theta)P(Y|Z, \Theta)}{P(Z|Y, \Theta^{(i)}) \log P(Y|\Theta^{(i)})} \right\}\end{aligned}$$

可以根据上面的不等式得到： $L(\Theta) \geq B(\Theta, \Theta^{(i)})$ ，且易证  $L(\Theta^{(i)}) = B(\Theta^{(i)}, \Theta^{(i)})$ ，因此我们找到了 $L(\Theta)$ 的下确界。只要通过调整参数 $\Theta$ 增大 $B(\Theta, \Theta^{(i)})$ 我们就能同时使 $L(\Theta)$ 增大。选择 $\Theta^{(i+1)} = \arg \max_{\Theta} B(\Theta, \Theta^{(i)})$ ：

$$\begin{aligned}\Theta^{(i+1)} &= \arg \max_{\Theta} B(\Theta, \Theta^{(i)}) \\ &= \arg \max_{\Theta} \left\{ L(\Theta^{(i)}) + \sum_z \left[ P(Z|Y, \Theta^{(i)}) \log \frac{P(Z|\Theta)P(Y|Z, \Theta)}{P(Z|Y, \Theta^{(i)}) \log P(Y|\Theta^{(i)})} \right] \right\} \\ &= \arg \max_{\Theta} \sum_z P(Z|Y, \Theta^{(i)}) \{ \log [P(Z|\Theta)P(Y|Z, \Theta)] - \log [P(Z|Y, \Theta^{(i)}) \log P(Y|\Theta^{(i)})] \} \\ &= \arg \max_{\Theta} \{ \sum_z P(Z|Y, \Theta^{(i)}) \log [P(Z|\Theta)P(Y|Z, \Theta)] \} \\ &= \arg \max_{\Theta} \{ \sum_z P(Z|Y, \Theta^{(i)}) \log P(Y, Z|\Theta) \} \\ &= \arg \max_{\Theta} Q(\Theta, \Theta^{(i)})\end{aligned}$$

因此我们就得到了Q函数。

