

## 7\_期望最大算法

不完全数据：观测随机变量 $Y$ 。

完全数据：观测随机变量 $Y$ 和隐随机变量 $Z$ 。

含有隐变量 $Z$ 的概率模型，目标是极大化观测变量 $Y$ 关于参数 $\theta$ 的对数似然函数，即

$$\max_{\theta} L(\theta)$$

其中，

$$\begin{aligned} L(\theta) &= \log P(Y|\theta) \\ &= \log \sum_Z P(Y, Z|\theta) \\ &= \log \left( \sum_Z P(Y|Z, \theta) P(Z|\theta) \right) \end{aligned}$$

对数似然函数 $L(\theta)$ 与第 $i$ 次迭代后的对数似然函数 $L(\theta^{(i)})$ 的差

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

令

$$B(\theta, \theta^{(i)}) = L(\theta^{(i)}) + \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y|Z, \theta) P(Z|\theta)}{P(Z|Y, \theta^{(i)}) P(Y|\theta^{(i)})}$$

则

$$L(\theta) \geq B(\theta, \theta^{(i)})$$

即函数 $B(\theta, \theta^{(i)})$ 是 $L(\theta)$ 的一个下界。

选择 $\theta^{(i+1)}$ 使 $B(\theta, \theta^{(i)})$ 达到极大，即

$$\begin{aligned} \theta^{(i+1)} &= \arg \max_{\theta} B(\theta, \theta^{(i)}) \\ &= \arg \max_{\theta} \left( L(\theta^{(i)}) + \sum_Z P(Z|Y, \theta^{(i)}) \log \frac{P(Y|Z, \theta) P(Z|\theta)}{P(Z|Y, \theta^{(i)}) P(Y|\theta^{(i)})} \right) \\ &= \arg \max_{\theta} \left( \sum_Z P(Z|Y, \theta^{(i)}) \log(P(Y|Z, \theta)) P(Z|\theta) \right) \\ &= \arg \max_{\theta} \left( \sum_Z P(Z|Y, \theta^{(i)}) \log P(Y, Z|\theta) \right) \end{aligned}$$

EM算法:

输入: 观测随机变量数据 $Y$ , 隐随机变量数据 $Z$ , 联合分布 $P(Y, Z|\theta)$ , 条件分布 $P(Y|Z, \theta)$ ;

输出: 模型参数 $\theta$

1. 初值 $\theta^{(0)}$

2.  $E$ 步:

$$\begin{aligned} Q(\theta, \theta^{(i)}) &= E_Z [\log P(Y, Z|\theta) | Y, \theta^{(i)}] \\ &= \sum_Z \log P(Y, Z|\theta) \cdot P(Z|Y, \theta^{(i)}) \end{aligned}$$

3.  $M$ 步:

$$\theta^{(i+1)} = \arg \max_{\theta} Q(\theta, \theta^{(i)})$$

4. 重复2. 3., 直到收敛。

$Q$ 函数: 完全数据的对数似然函数 $\log P(Y, Z|\theta)$ 关于在给定观测数据 $Y$ 和当前参数 $\theta_{(i)}$ 下对未观测数据 $Z$ 的条件概率分布 $P(Z|Y, \theta_{(i)})$ 的期望

$$Q(\theta, \theta_{(i)}) = E_Z [\log P(Y, Z|\theta) | Y, \theta_{(i)}]$$