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。
对应HMM的观测序列O 状态序列I 对应HMM的λ
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θ的对数似然函数，即

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

其中，

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

HMM的问题2，给观测序列，求模型λ，使max P(O|λ)

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

$$\begin{aligned} L(\theta) - L\left(\theta^{(i)}\right) &= \log\left(\sum_Z P(Y|Z,\theta)\,P(Z|\theta)\right) - \log P\left(Y|\theta^{(i)}\right) \\ &= \log\left(\sum_Z P\left(Z|Y,\theta^{(i)}\right) \frac{P(Y|Z,\theta)\,P(Z|\theta)}{P\left(Z|Y,\theta^{(i)}\right)}\right) - \log P\left(Y|\theta^{(i)}\right) \\ &\geq \sum_Z P\left(Z|Y,\theta^{(i)}\right) \log \frac{P(Y|Z,\theta)\,P(Z|\theta)}{P\left(Z|Y,\theta^{(i)}\right)} - \log P\left(Y|\theta^{(i)}\right) \\ &= \underbrace{\boxed{\sum_Z P\left(Z|Y,\theta^{(i)}\right)}}_{1} \log \frac{P(Y|Z,\theta)\,P(Z|\theta)}{P\left(Z|Y,\theta^{(i)}\right)\,P\left(Y|\theta^{(i)}\right)} \end{aligned}$$

分子分母同乘

jensen不等式

Jensen不等式：

$$\log \sum_j \lambda_j y_j \geq \sum_j \lambda_j \log y_j$$

令

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

则

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

即函数$B\left(\theta,\theta^{(i)}\right)$ 是$L(\theta)$ 的一个下界。
选择$\theta^{(i+1)}$使$B\left(\theta,\theta^{(i)}\right)$达到极大，即

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

这里theta(i)已知，所以第一项忽略，分母也忽略

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$P\left(Y \mid Z, \theta\right)$；
输出：模型参数θ

1. 初值$\theta^{(0)}$
2. E步：
$$Q\left(\theta, \theta^{(i)}\right) = E_Z\left[\log P(Y,Z|\theta)|Y, \theta^{(i)}\right] \\ = \sum_Z \log P(Y,Z|\theta) \cdot P\left(Z|Y, \theta^{(i)}\right)$$
3. M步：
$$\theta^{(i+1)} = \arg\max_\theta Q\left(\theta, \theta^{(i)}\right)$$
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$P\left(Z|Y, \theta_{(i)}\right)$的期望
$$Q\left(\theta, \theta_{(i)}\right) = E_Z\left[\log P(Y,Z|\theta)|Y, \theta_{(i)}\right]$$