EM算法

为什么需要EM算法：

概率模型有时既含有观测变量，又含有隐变量或者潜在变量。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用极大似然估计法，或者贝叶斯估计法估计模型参数。但是，当模型含有隐变量时，就不能简单地使用这些估计方法。

EM算法就是含有隐变量的概率模型参数的极大似然估计方法。

三硬币模型

假设有3枚硬币，分别记作A，B，C，这些硬币正面出现的概率分别是π，p和q。

进行如下硬币试验：

先掷硬币A，根据其结果选出硬币B或硬币C，正面选硬币B，反面选硬币C；然后掷选出的硬币，掷硬币的结果，出现正面记作1，出现反面记作0；独立地重复n次实验，n=10，观测结果如下：1，1，0，1，0，0，1，0，1，1

假设只能观测到掷硬币的结果，不能观测掷硬币的过程，问如何估计三硬币正面出现的概率，即三硬币模型的参数。

解：对每一次试验可以进行如下建模

$$P(y|\theta)=\sum_z P(y,z|\theta)=\sum_z P(z|\theta)P(y|z,\theta)$$

$$=P(z=1|\theta)P(y|z=1,\theta)+P(z=0|\theta)P(y|z=0,\theta)$$

$$=\begin{cases}\pi p+(1-\pi)q, & \text{if } y=1\\ \pi(1-p)+(1-\pi)(1-q) & \text{if } y=0;\end{cases}$$

$$=\pi p^y(1-p)^{1-y}+(1-\pi)q^y(1-q)^{1-y}$$

其中，随机变量y是观测变量，表示一次试验观测的结果是1或0；随机变量z是隐变量，表示未观测到的掷硬币A的结果；$\theta=(\pi,p,q)$是模型参数。将观测数据表示为

$Y=(Y_1,Y_2,\cdots,Y_n)^T$，未观测数据表示为$Z=(Z_1,Z_2,\cdots Z_n)^T$。

观测数据的似然函数为：

$$P(Y|\theta)=\sum_z P(Z|\theta)P(Y|Z,\theta)=\prod_{j=1}^{n}P(y_j|\theta)$$

$$=\prod_{j=1}^{n}\left[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]$$

求模型参数$\theta=(\pi,p,q)$的极大似然估计，即使用对数似然函数来进行参数估计可得：

$$\hat{\theta}=\arg\max_{\theta}\ln P(Y|\theta)$$

$$=\arg\max_{\theta}\ln\prod_{j=1}^{n}\left[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]$$

$$=\arg\max_{\theta}\sum_{j=1}^{n}\ln\left[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}\right]$$

EM的导出

Jensen(琴生)不等式:

若f是凸函数,则:

$$f(tx_1+(1-t)x_2)\leq tf(x_1)+(1-t)f(x_2)$$

其中$t\in[0,1]$. 同理,若f是凹函数,则只需将上式中的$\leq$换成$\geq$即可

将上式中的t推广到n个同样成立,也即

$$f(t_1x_1+t_2x_2+\dots+t_nx_n)\leq t_1f(x_1)+t_2f(x_2)+\dots+t_nf(x_n)$$

其中,$t_1,t_2,\dots t_n\in[0,1]$。同理,若f是凹函数,则只需将上式中的$\leq$换成$\geq$即可。

将上式中的t推广到n个同样成立,也即:

$$f(t_1x_1+t_2x_2+\dots t_nx_n)\leq t_1f(x_1)+t_2f(x_2)+\dots+t_nf(x_n)$$

其中,$t_1,t_2,\dots,t_n\in[0,1]$, $t_1+t_2+\dots+t_n=1$. 在概率论中常以以下形式出现:

$$\varphi(E[X])\leq E[\varphi(X)]$$

其中,X是随机变量,φ是凸函数,E[X]表示X的期望

$$L(\theta)=\ln P(Y|\theta)=\ln\sum_Z P(Y,Z|\theta)=\ln\left(\sum_Z P(Y|Z,\theta)P(Z|\theta)\right)$$

$$L(\theta)-L(\theta^{(i)})=\ln\left(\sum_Z P(Y|Z,\theta)P(Z|\theta)\right)-\ln P(Y|\theta^{(i)})$$

$$=\ln\left(\sum_Z P(Z|Y,\theta^{(i)})\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})}\right)-\ln P(Y|\theta^{(i)})$$

套用琴生不等式可得:

$$\geq\sum_Z P(Z|Y,\theta^{(i)})\ln\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})}-\ln P(Y|\theta^{(i)})$$

$$L(\theta)-L(\theta^{(i)})\geq\sum_Z P(Z|Y,\theta^{(i)})\ln\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})}-\sum_Z P(Z|Y,\theta^{(i)})\ln P(Y|\theta^{(i)})$$

$$=\sum_Z P(Z|Y,\theta^{(i)})\left(\ln\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})}-\ln P(Y|\theta^{(i)})\right)$$

$$=\sum_Z P(Z|Y,\theta^{(i)})\ln\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})P(Y|\theta^{(i)})}$$

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

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

$$B(\theta^{(i+1)},\theta^{(i)})\geq B(\theta^{(i)},\theta^{(i)})$$

由于$B(\theta^{(i)},\theta^{(i)})=L(\theta^{(i)})$,所以可以进一步推得:

$$L(\theta^{(i+1)})\geq B(\theta^{(i+1)},\theta^{(i)})\geq B(\theta^{(i)},\theta^{(i)})=L(\theta^{(i)})$$

NO. 5
DATE

因此，任何可以使 $B(\theta,\theta^{(i)})$ 增大的 θ，也可以使 $L(\theta)$ 增大，于是问题转化为了求解能使得 $B(\theta,\theta^{(i)})$ 达到极大的 $\theta^{(i+1)}$，即

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

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

$$= \arg\max_{\theta}\left(\sum_Z P(Z|Y,\theta^{(i)}) \ln P(Y|Z,\theta)P(Z|\theta)\right)$$

$$= \arg\max_{\theta}\left(\sum_Z P(Z|Y,\theta^{(i)}) \ln P(Y,Z|\theta)\right)$$

$$= \arg\max_{\theta} Q(\theta,\theta^{(i)}) \qquad 式[7.3]$$