高斯混合模型(GMM)

1. 模型

$y_1, y_2, \cdots, y_K,$

$P(y|\theta)=\sum_{i=1}^{K}\alpha_k\Phi(y|\theta_k)$　$\forall k\geq 0, \Sigma_k\alpha_k=1$, $\theta_k=(\mu_k,\sigma_k^2)$, $\Phi(y|\theta_k)=\frac{1}{\sqrt{2\pi}\sigma_k}\exp(-\frac{(y-\mu_k)^2}{2\sigma_k^2})$

$y_j \to$ ① 依据 α_k 选择第k个高斯分布 $\Phi(y|\theta_j)$

$\to$ ② 依据 $\Phi(y|\theta_j)$ 选择 y_j

- y_j 是已知的 $j=1,\cdots,N$

隐变量 r_{jk} $\begin{cases} 1 & \text{第j个观测来自k} \\ 0 & \text{otherwise} \end{cases}$

不可观测的

2. 参数估计

2.1 EM算法

如果模型有观测数据(observed), 隐变量(latent)

例(三硬币模型)

A　B　C

π　p　q

$A\begin{cases} 正 & B \\ 反 & C \end{cases}$

Result: 1, 1, 0, 1, 0, 0, 1, 0, 1, 1

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

$$=\sum_z P(z|\theta)P(y|z,\theta)=\pi\cdot p^y(1-p)^{1-y}+(1-\pi)q^y(1-q)^{1-y}$$

$$P(Y|\theta)=\prod_{j=1}^{n}[\pi p^{y_j}(1-p)^{1-y_j}+(1-\pi)q^{y_j}(1-q)^{1-y_j}]$$

一般用EM算法求解

$y^{(t)}=(\pi^{(t)}, p^{(t)}, q^{(t)})$

E步：计算在给定模型参数的情形下，y_j来自B的概率

Expectation $P(z_j=1|y_j,\theta)=\frac{P(z_j,y_j|\theta)}{\sum_{z_j}P(z_j,y_j|\theta)}$

$$\mu^{(i+1)}=E(z_j)=\frac{\pi^{(i)}(p^{(i)})^{y_j}(1-p^{(i)})^{1-y_j}}{\pi^{(i)}(p^{(i)})^{y_j}(1-p^{(i)})^{1-y_j}+(1-\pi^{(i)})(q^{(i)})^{y_j}(1-q^{(i)})^{1-y_j}}$$

M步：$P(Y,Z|\theta)=\prod_{j=1}^{n}[\pi p^{y_j}(1-p)^{1-y_j}]^{z_j}[(1-\pi)q^{y_j}(1-q)^{1-y_j}]^{1-z_j}$

Maximize

$$\pi^{(i+1)}=\frac{1}{n}\sum_j\mu_j^{(i+1)}$$

$$p^{(i+1)}=\frac{\sum_j\mu_j^{(i+1)}y_j}{\sum_j\mu_j^{(i+1)}},\quad q^{(i+1)}=\frac{\sum_j(1-\mu_j^{(i+1)})y_j}{\sum_j(1-\mu_j^{(i+1)})}$$

- Y：观测数据
  Z：隐变量 > 完全数据 complete data

EM算法：输入 Y，$P(Y,Z|\theta)$ $P(Z|\theta)$
输出 θ

PseudoCode：(1) 选择初始值$\theta^{(0)}$

(2) E步：$\theta^{(i)}$是第i步的估计值

求条件概率 $P(Z|\theta,Y)=\mu^{(i)}$

(3) M步：$Q(\theta,\theta^{(i)})=E_Z\{\log(P(Y,Z|\theta))|Y,\theta^{(i)}\}$

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

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

2.2 EM算法的导出

① 为什么EM算法可以实现对数据的极大似然估计？

似然函数：$L(\theta)=\log P(Y|\theta)=\log\{\sum_Z P(Y,Z|\theta)\}$

$$=\log\{\sum_Z P(Y|Z,\theta)P(Z|\theta)\}$$

$\theta^{(i)}$ 希望 θ 增加 $L(\theta)$

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

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

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

$\log\sum\lambda_j y_j\geq\sum\lambda_j\log y_j$

if $\lambda_j\geq 0$ $\sum_j\lambda_j=1$

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

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

$\therefore L(\theta)\geq\underbrace{B(\theta,\theta^{(i)})}_{\text{下界}}$，选择 $\theta^{(i+1)}=\arg\max_\theta B(\theta,\theta^{(i)})$

$$\therefore \theta^{(i+1)}=\arg\max_\theta\{L(\theta^{(i)})+\sum_Z P(Z|Y,\theta^{(i)})\log\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})P(Y|\theta^{(i)})}\}$$

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

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

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

(1) EM对初值敏感

(2) 在迭代中使 $L(\theta)\uparrow$

(3) 在一定条件下可以保证收敛，但不一定收敛到极大值点

2.3 通过EM算法求解GMM

(1) 明确隐变量，写出完全数据的对数似然 $(y_j, r_{j1}, r_{j2}, \cdots, r_{jK})\ j=1,2,\cdots,N$

$$P(y,r|\theta) = \prod_{j=1}^{N} P(y_j, r_{j1}, r_{j2}, \cdots, r_{jK}|\theta)$$

$$= \prod_{j=1}^{N} \prod_{k=1}^{K} [\alpha_k \Phi(y_k|\theta_k)]^{r_{jk}}$$

$$= \prod_{k=1}^{K} \alpha_k^{r_k} \prod_{j=1}^{N} [\Phi(y_k|\theta_k)]^{r_{jk}}, \quad r_k = \sum_j r_{jk}$$

$$= \prod_{k=1}^{K} \alpha_k^{r_k} \prod_{j=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right)\right]^{r_{jk}}$$

$$\log P(y,r|\theta) = \sum_{k=1}^{K} \left\{ r_k \log \alpha_k + \sum_{j=1}^{N} r_{jk} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \log \sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]\right.$$

$$\hat{r}_{jk} = E(r_{jk}|y,\theta) = P(r_{jk}=1|y_{jk},\theta)$$ ← 子模型k对于y_j的响应度

$$= \frac{P(r_{jk}=1, y|\theta)}{\sum_{k=1}^{K} P(r_{jk}, y_j|\theta)} = \frac{P(y|r_{jk}=1,\theta) P(r_{jk}=1|\theta)}{\sum_k P(y_j|r_{jk}=1,\theta) P(r_{jk}=1|\theta)}$$

$$= \frac{\alpha_k \cdot \Phi(y_j|\theta_k)}{\sum_k \alpha_k \cdot \Phi(y_j|\theta_k)}$$

$$Q(\theta,\theta^{(i)}) = \sum_{k=1}^{K} \left\{ \underbrace{\sum_{j=1}^{N} \hat{r}_{jk}}_{= r_k} \log \alpha_k + \sum_{j=1}^{N} \hat{r}_{jk} \left[\log \frac{1}{\sqrt{2\pi}} - \log \sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]\right\}$$

$$\theta^{(i+1)} = \arg\max_{\theta} Q(\theta,\theta^{(i)}), \quad s.t. \sum_{k=1}^{K} \alpha_k = 1$$

$$\Rightarrow \hat{\mu}_k = \frac{\sum_{j=1}^{N} \hat{r}_{jk} y_j}{\sum_{j=1}^{N} \hat{r}_{jk}}, \quad \hat{\sigma}_k^2 = \frac{\sum_j \hat{r}_{jk} (y_j-\mu_k)^2}{\sum_j \hat{r}_{jk}}, \quad \hat{\alpha}_k = \frac{\sum_{j=1}^{N} \hat{r}_{jk}}{N}$$

GMM的EM算法：

(1) 选择初值开始迭代

(2) 求响应度 $\hat{r}_{jk} = \dfrac{\alpha_k^{(i)} \Phi(y_j|\theta_k^{(i)})}{\sum_{k=1}^{K} \alpha_k^{(i)} \Phi(y_j|\theta_k^{(i)})}$

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

2.4 EM算法的收敛性

定理9.1：设 $P(Y|\theta)$ 为观测数据的似然函数，$\theta^{(i)}(i=1,2,\cdots)$ 为EM算法得到的参数估计序列，$P(Y|\theta^{(i)})(i=1,2,\cdots)$ 为对应的似然函数序列，则 $P(Y|\theta^{(i)})$ 是单调递增的，即 $P(Y|\theta^{(i+1)}) \geq P(Y|\theta^{(i)})$

证明：$\because P(Y|\theta)=\dfrac{P(Y,Z|\theta)}{P(Z|Y,\theta)}$

$$\log P(Y|\theta)=\log P(Y,Z|\theta)-\log P(Z|Y,\theta)$$

$$Q(\theta,\theta^{(i)})=\sum_{Z}\log P(Y,Z|\theta)\cdot P(Z|Y,\theta^{(i)})$$

令 $H(\theta,\theta^{(i)})=\sum_{Z}\log P(Z|Y,\theta)\cdot P(Z|Y,\theta^{(i)})$

$$\therefore \log P(Y|\theta)=Q(\theta,\theta^{(i)})-H(\theta,\theta^{(i)})$$

$$\therefore \log P(Y|\theta^{(i+1)})-\log P(Y|\theta^{(i)})=\left[Q(\theta^{(i+1)},\theta^{(i)})-Q(\theta^{(i)},\theta^{(i)})\right]-\left[H(\theta^{(i+1)},\theta^{(i)})-H(\theta^{(i)},\theta^{(i)})\right]$$

又 $\because$
$$H(\theta^{(i+1)},\theta^{(i)})-H(\theta^{(i)},\theta^{(i)})=\sum_{Z}\left(\log\frac{P(Z|Y,\theta^{(i+1)})}{P(Z|Y,\theta^{(i)})}\cdot P(Z|Y,\theta^{(i)})\right)$$
$$\leq \log\left(\sum_{Z}\frac{P(Z|Y,\theta^{(i+1)})}{P(Z|Y,\theta^{(i)})}\cdot P(Z|Y,\theta^{(i)})\right)$$
$$=0$$

$$Q(\theta^{(i+1)},\theta^{(i)})=\max Q(\theta,\theta^{(i)})\geq Q(\theta^{(i)},\theta^{(i)})$$

$$\therefore \log P(Y|\theta^{(i+1)})-\log P(Y|\theta^{(i)})\geq 0$$

$P(Y|\theta^{(i+1)})\geq P(Y|\theta^{(i)})$ 才保证了收敛性