第9章 EM算法及其推广

9.1 EM算法的引入

用来估计概率分布，但观测数据有缺失

例 9.1 三硬币模型

不完全数据

完全数据

$z \sim b(1, \pi)$

↓ (z_1, y_1)

我们观测不到的数据

若 $z_1 = 1$, $y_1 \sim b(1, p)$

若 $z_1 = 0$, $y_1 \sim b(1, q)$

能看的只有 y_1，

不能看到 z_1 的值，只知道 z 服一个这样的分布，并且 π, p, q 都不知道。

(z_1, y_1) (z_2, y_2) $\cdots$ (z_N, y_N) $\theta = (\pi, p, q)^T$

(z, y) 完全数据（不能看到）→ $P(y, z) = P(z) \cdot P(y|z)$

y 不完全数据（能看到）— 观测值

我们想估计的是 $\theta = (\pi, p, q)^T$，一般用极大似然估计

极大似然要写出给定 θ 情况下与观测值的联合概率密度

$$P(y|\theta) = \sum_z P(\underbrace{y, z|\theta}_{\text{边缘分布}}) = \pi p^{y}(1-p)^{1-y} + (1-\pi) q^{y} (1-q)^{1-y}$$

整理

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

∵每个 y_i 独立，∴有

$$\prod_{i=1}^{N} P(y_i|\theta) = \prod \left[\sum_z P(y_i, z|\theta) \right] = \prod [\quad]$$

∵上式特别复杂，若要用极大似然法去求，则要最大化似然函数，对其求p.q的导不现实。∴用EM算法（迭代）

★ EM不直接处理上面~~式子~~似然函数，要：

将最大化<u>观测数据</u>的似然函数 转化为

最大化<u>完全数据</u>的似然函数

书上有2种解释 $\begin{cases} \text{导出} \\ \text{推广} \end{cases}$

$\Rightarrow$ $p(y, z \mid \theta)$ $\qquad \ln \prod_{i=1}^{n} p(y_i, z_i \mid \theta)$

（我们不知道z_i的数据）

∴ EM算法中的E步解决了这个问题

E：把与z_i有关的项用 $E(z)$ 去替换 $\qquad E_i \to E(z) \to \theta, y_i$

对不同的i来说，E(z)是不变的

如何求E(z)呢？

∵ z是个随机变量，∴求期望时就要知道z的分布（概率密度）

但这个分布取决于θ与y_i，用迭代法：

1) 给θ 1个初值 与y_i一并用来求z的期望 $\theta^{(i)}$ ← 第i步给的θ

2) 将期望代入似然函数，再最大化似然函数

M：$\theta^{(i+1)} = \arg\max_{\theta} \ln \prod_{i=1}^{n} (p_i, E(z) \mid \theta)$

求得$\theta^{(i+1)}$

不断迭代

EM算法的导出

(待估参数) θ . Y (观测向量) Z (隐)

$$L(\theta) = \ln P(Y|\theta) \quad —①$$

$$= \ln \sum_{Z} P(Y, Z|\theta) —②$$

$$= \ln \left[\sum_{Z} P(Z|\theta) P(Y|Z,\theta)\right] —③$$

①= ②的边缘密度之和

②→③ 依据乘法公式

$\theta^{(i)}$ ← 第i步求出的θ

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

分析： $\ln P_1 P_2 \cdots P_N$ 好处理

但 $\ln \sum P_1 P_2$ 里多了求和号不好处理

∴ 用 Jessan 不等式 $f(\sum_i \alpha_i x_i) \geq \sum_i \alpha_i f(x_i)$

f为凸函数 (0,1)

要想办法凑出 α_i ∵是对Z求和. 概率求和为1.

∴ 构造一个Z的分布

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

$$\geq \sum_{Z} P[Z|Y,\theta^{(i)}] \ln \frac{P(Z|\theta) P(Y|Z,\theta)}{P(Z|Y,\theta^{(i)})} - \sum_{Z} P(Z|Y,\theta^{(i)}) \cdot \underline{\ln P(Y|\theta^{(i)})}$$

人为加上. ∵是对Z求和. ∴此项为1

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

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

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

最大化 $L(\theta)$ 即最大化 $B(\theta, \theta^{(i)})$

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

去掉与θ无关的

$$= \arg\max_{\theta} \sum_{Z} P(Z|Y, \theta^{(i)}) \cdot (\ln P(Z|\theta) \cdot P(Y|Z, \theta))$$

$$= \arg\max_{\theta} \underbrace{\sum_{Z} P(Z|Y, \theta^{(i)}) \ln P(Y, Z|\theta)}_{Q(\theta, \theta^{(i)})}$$

即 $\arg\max Q(\theta, \theta^{(i)})$ ←M步

EM算法及其推广(如何用EM算法求解混合高斯模型中的参数)

高斯混合模型

$$P(y|\theta)=\sum_{k=1}^{K}\alpha_k\phi(y|\theta_k)$$

→ 第k个高斯模型的密度函数

→ α_k 表示y以 α_k 的概率来自于第k个概率分布

待估参数

$\theta=(\alpha_1,\cdots,\alpha_k,u_1,\sigma_1^2,\cdots,u_k,\sigma_k^2)$

$\gamma_1=(\gamma_{11},\gamma_{12},\cdots,\gamma_{1k})$

$$\begin{array}{c} \uparrow 取 \\ 隐变量 \end{array}\left\{\begin{array}{l}(1,0,\cdots\cdots,0)\\(0,1,\cdots\cdots,0)\\ \vdots \\(0,0,\cdots\cdots,1)\end{array}\right\}\begin{array}{c}的概\\率为\end{array}\left\}\begin{array}{l}\alpha_1\\\alpha_2\\\vdots\\\alpha_k\end{array}\right.$$

(表示当前的y来自哪个高斯分布

$\therefore P(\gamma_1,y_1|\theta)=P(\gamma_1|\theta)\cdot P(y_1|\gamma_1,\theta)$

$$=\alpha_1^{\gamma_{11}}\cdot\alpha_2^{\gamma_{12}}\cdots\alpha_k^{\gamma_{1k}}\cdot\phi(y_1|\theta_1)^{\gamma_{11}}\phi(y_1|\theta_2)^{\gamma_{12}}\cdots\phi(y_1|\theta_k)^{\gamma_{1k}}$$

$$=\prod_{k=1}^{K}[\alpha_k\phi(y_1|\theta_k)]^{\gamma_{1k}}$$

表示第1个样本点的完全数据的密度函数

∴对所有样本点来说

$$P(y,\gamma|\theta)=\prod_{j=1}^{N}\prod_{k=1}^{K}[\alpha_k\phi(y_j|\theta_k)]^{\gamma_{jk}} \quad —①$$

$\cancel{\alpha_k^{\gamma_{jk}}}$ $\prod_{k=1}^{K}\prod_{j=1}^{N}\alpha_k^{\gamma_{jk}}$

$$=\alpha_k^{\sum_{j=1}^{N}\gamma_{jk}}$$

(表示在N个样本点中,共有多少个是来自第k个高斯分布) 用 n_k 表示 $(n_1+n_2+\cdots n_k=N)$

①式 → $\prod_{k=1}^{K}\alpha_k^{n_k}\prod_{j=1}^{N}[\phi(y_j|\theta_k)]^{\gamma_{jk}}$

将$\varphi(y_j|\theta_k)$展开得

$$p(y,r|\theta)=\prod_{k=1}^{K}\alpha_k^{n_k}\prod_{j=1}^{N}\left[\frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right)\right]^{\gamma_{jk}}$$

EM对完全数据的对数函数求解

$$\ln p(y,r|\theta)=\sum_{k=1}^{K}\left\{n_k\ln\alpha_k+\sum_{j=1}^{N}\gamma_{jk}\left[\ln\frac{1}{\sqrt{2\pi}}-\ln\sigma_k-\frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]\right.$$ ③

E步：把上式与$\underbrace{\text{隐变量}}_{\gamma_{jk}}$有关的项都换成其期望

$$En_k=E\left[\sum_j\gamma_{jk}\right]=\sum_j E(\gamma_{jk})$$

$E(\gamma_{jk})=P(\gamma_{jk}=1)$ （γ_{jk}要么为1，要么为0）

求期望时，要根据上一步的θ^i与所有观测数据y、z来求

$$E(\gamma_{jk}|\theta^{(i)},y)=\underline{P(\gamma_{jk}=1|\theta^{(i)},y)}$$ 用贝叶斯公式求

$$=\frac{P(\gamma_{jk}=1,y_j|\theta)}{p(y_j|\theta)}$$ → 用全概率公式展开

$$=\frac{P(\gamma_{jk}=1,y_j|\theta)}{\sum_{k=1}^{K}P(\gamma_{jk}=1,y_j|\theta)}$$ → 用乘法公式

$$=\frac{\overbrace{P(\gamma_{jk}=1|\theta)}^{\alpha_k}\ \overbrace{P(y_j|\gamma_{jk}=1,\theta)}^{\varphi(y_j|\theta_k)}}{\sum_k P(y_j|\gamma_{jk}=1,\theta)\,P(\gamma_{jk}=1|\theta)}$$

$$=\frac{\alpha_k^{(i)}\cdot\varphi(y_j|\theta_k^{(i)})}{\sum\alpha_k^{(i)}\varphi(y_j|\theta_k^{(i)})}$$

记为 z_k

$$E(\gamma_{jk}\mid y,\theta^{(i)}) = \frac{\alpha_k^{(i)}\varphi(y_j\mid\theta_k^{(i)})}{\sum_k \alpha_k^{(i)}\varphi(y_j\mid\theta_k^{(i)})}$$

$\alpha_k^{(i)}$ $\theta_k^{(i)}$

代入③式中的 n_k 与 γ_{jk} 得 $\underline{Q(\theta,\theta^{(i)})}$

将 z_k 代入上面③式中得

~~$P(y,\gamma|\theta)$~~

$$\sum_{k=1}^{K}(N z_k)\cdot\ln\alpha_k + z_k\sum_{j=1}^{N}\left[\ln\frac{1}{\sqrt{2\pi}} - \ln\sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]$$

即

$$Q(\theta,\theta^{(i)}) = E_z \ln P(y,\gamma\mid\theta)$$

M步：要最大化Q函数，对每个θ的分量求导，让其等于0

要估计(γ_k, σ_k, μ_k)

$$\frac{\partial Q(\theta,\theta^{(i)})}{\partial \mu_k} = 0 \qquad \mu_k^{(i+1)} = \cdots$$

$$\frac{\partial Q(\theta,\theta^{(i)})}{\partial \sigma_k^2} = 0 \qquad \sigma_k^{2\,(k+1)} = \cdots$$

$$\begin{cases}\dfrac{\partial Q(\theta,\theta^{(i)})}{\partial \alpha_k} \\ \sum \alpha_k = 1\end{cases} \qquad \alpha_k^{(i+1)} = \cdots$$

书上 9.30~9.32 式