4.21 《机器学习基础》第7次作业

王宇哲 1800011828

T1

解. 对于二项 logistic 回归模型, 有

$$\log L(\theta)=\sum_{i=1}^{N}\left[y_i(w\cdot x_i+b)-\log(1+e^{w\cdot x_i+b})\right]$$

参数 θ 的极大似然估计

$$\hat{\theta}=(\hat{w},\hat{b})=\underset{\theta=(w,b)}{\arg\max}\log L(\theta)=\underset{\theta=(w,b)}{\arg\min}\left[-\log L(\theta)\right]$$

利用拟牛顿法求解, 算法框架如下:
(BFGS 算法)

输入: 目标函数 $f(\theta)=-\log L(\theta)$

$g(\theta)=\nabla f(\theta)$

精度要求 ε

输出: 极小点 $\hat{\theta}=\underset{\theta=(w,b)}{\arg\min}f(\theta)$

i) 任意选定初始 $\theta^{(0)}$, 取正定对称矩阵 B_0, 令 $k=0$

ii) 计算 $g_k=g(\theta^{(k)})$

若 $\|g_k\|<\varepsilon$, 输出 $\hat{\theta}=\theta^{(k)}$, 停止;
否则转到 iii)

iii) 由

$$B_k p_k=-g_k$$

求出 p_k

iv) 一维搜索: 求 λ_k s.t.

$$f(\theta^{(k)}+\lambda_k p_k)=\min_{\lambda\ge 0}f(\theta^{(k)}+\lambda p_k)$$

v) 令 $\theta^{(k+1)}=\theta^{(k)}+\lambda_k p_k$

vi) 计算 $g_{k+1}=g(\theta^{(k+1)})$

若 $\|g_{k+1}\|<\varepsilon$, 输出 $\hat{\theta}=\theta^{(k+1)}$, 停止;
否则, 计算

$$B_{k+1}=B_k+\frac{y_k y_k^T}{y_k^T\delta_k}-\frac{B_k\delta_k\delta_k^T B_k}{\delta_k^T B_k\delta_k}$$

其中

$$y_k=g_{k+1}-g_k$$
$$\delta_k=\theta^{(k+1)}-\theta^{(k)}$$

vii) 令 $k=k+1$, 转到 iii)

T2

证: 1)
极大似然估计:

先验概率 $$\hat{P}(Y=c_k)=\frac{\sum_{j=1}^{N}I(y_j=c_k)}{N} \quad ①$$

$k=1,2,\cdots,K$

条件概率

$$\hat{P}(X^{(i)}=a_l^{(i)}|Y=c_k)=\frac{\sum_{j=1}^{N}I(x_j^{(i)}=a_l^{(i)},y_j=c_k)}{\sum_{j=1}^{N}I(y_j=c_k)} \quad ②$$

$l=1,2,\cdots,m_i,\ i=1,2,\cdots,n$

先证①. 记

$$p=P(Y=c_k)$$
$$q=\sum_{j=1}^{N}I(y_j=c_k)$$

似然函数

$$L(p)=C_N^q p^q(1-p)^{N-q}$$

令 $\frac{\partial L(p)}{\partial p}=0$, 即

$$qp^{q-1}(1-p)^{N-q}=(N-q)(1-p)^{N-q-1}p^q$$

非平凡解为 $\hat{p}=\frac{q}{N}=\frac{\sum_{j=1}^{N}I(y_j=c_k)}{N}$

即 $\hat{P}(Y=c_k)=\frac{\sum_{j=1}^{N}I(y_j=c_k)}{N}$；

类似地，对②，记

$$p_l=P(X^{(i)}=a_l^{(i)}\mid Y=c_k)$$
$$q_l=\sum_{j=1}^{N}I(x_j^{(i)}=a_l^{(i)},\ y_j=c_k)$$

似然函数

$$L(p_l)=C_q^{q_l}\ p_l^{q_l}(1-p_l)^{q-q_l}$$

经过相似计算过程，即得

$$\hat{P}(X^{(i)}=a_l^{(i)}\mid Y=c_k)=\hat{p}_l=\frac{\sum_{j=1}^{N}I(x_j^{(i)}=a_l^{(i)},y_j=c_k)}{\sum_{j=1}^{N}I(y_j=c_k)}\ ;$$

2）贝叶斯估计：

先验概率

$$\hat{P}_\lambda(Y=c_k)=\frac{\sum_{j=1}^{N}I(y_j=c_k)+\lambda}{N+K\lambda}\quad ③$$

$k=1,2,\cdots,K$

条件概率

$$\hat{P}_\lambda(X^{(i)}=a_l^{(i)}\mid Y=c_k)=\frac{\sum_{j=1}^{N}I(x_j^{(i)}=a_l^{(i)},y_j=c_k)+\lambda}{\sum_{j=1}^{N}I(y_j=c_k)+m_i\lambda}\quad ④$$

$l=1,2,\cdots,m_i$
$i=1,2,\cdots,n$

先证③.

考虑 $P_\lambda(Y=c_i)=\theta_i$，$i=1,\cdots,K$

为随机变量

c_i 分布为多项分布. 故共轭先验分布为 Dirichlet 分布. 记参数为 λ，则有

$$p(\theta_1,\cdots,\theta_K;\lambda)=\frac{1}{B(\lambda)}\prod_{i=1}^{K}\theta_i^{\lambda-1}\quad ⑤$$

训练集 $D=\{(x_i,y_i)\}_{i=1}^{N}$

记 $M_i=\sum_{j=1}^{N}I(y_j=c_i)$

$i=1,2,\cdots,K$

令 $\theta=(\theta_1,\cdots,\theta_K)$

$M=(M_1,\cdots,M_K)$

则后验分布

$$p(\theta\mid M)=\frac{p(M\mid\theta)\,p(\theta)}{\int p(M\mid\theta)p(\theta)d\theta}$$

考虑后验概率 $p(M\mid\theta)$ 服从多项分布.

$$p(M\mid\theta)=\frac{K!}{M_1!\cdots M_K!}\theta_1^{M_1}\cdots\theta_K^{M_K}\quad ⑥$$

由⑤⑥，又由 $\int p(M\mid\theta)p(\theta)d\theta$ 为定值，知

$$p(\theta\mid M)\propto\prod_{i=1}^{K}\theta_i^{M_i+\lambda-1}$$

也服从 Dirichlet 分布

因此

$$\hat{P}_\lambda(Y=c_k)=E(\theta_k)=\frac{M_k+\lambda}{N+K\lambda}=\frac{\sum_{j=1}^{N}I(y_j=c_k)+\lambda}{N+K\lambda}\ ;$$

④的证明是完全类似的，只需考虑

$$P_\lambda(\cancel{Y=c_i,}X^{(i)}=a_l^{(i)}\mid Y=c_k)=\theta_l\ ,$$

$l=1,2,\cdots,m_i$

$$M_l = \sum_{j=1}^{N} I(x_j^{(i)} = a_l^{(i)}, y_j = c_k)$$

使用完全相同的证明过程，即得

$$\hat{P}_\lambda(X^{(i)} = a_l^{(i)} | Y = c_k) = E(\theta_l)$$
$$= \frac{\sum_{j=1}^{N} I(x_j^{(i)} = a_l^{(i)}, y_j = c_k) + \lambda}{\sum_{j=1}^{N} I(y_j = c_k) + m_i \lambda}$$

证毕！