# 大数据机器学习

# 第一次作业

刘培源 2023214278

1. 说明伯努利模型的极大似然估计以及贝叶斯估计中的统计学习方法三要素。伯努利模型是定义在取值为 0 与 1 的随机变量上的概率分布。假设观测到伯努利模型 n 次独立的数据生成结果，其中 k 次的结果为 1，这时可以用极大似然估计或贝叶斯估计来估计结果为 1 的概率。

伯努利模型的概率分布函数为 $f_p(X=x)=p^x(1-p)^{(1-x)}$，

其中 $p$ 为随机变量 $X$ 取 1 的概率。

统计学习方法三要素为模型、策略、算法

由上面两点，可以得出：

1、极大似然估计：

模型：$\mathcal{F}=\{f \mid f_p(X=x)=p^x(1-p)^{(1-x)}\}$

策略：极大似然估计

算法：$\arg\max\limits_{p} -L(f_p(X=x))$

2、贝叶斯估计：

模型：$\mathcal{F}=\{f \mid f_p(X=x)=p^x(1-p)^{(1-x)}\}$

策略：求后验分布，然后计算期望

算法：$E_\pi[p \mid A_1,\dots,A_n]=\int_0^1 p\times\pi(p \mid A_1,\dots,A_n)\,dp$

$$=\int p\,\frac{f(A_1,\dots A_n \mid p)\,\pi(p)}{\int f(A_1,\dots,A_n \mid p)\,\pi(p)\,dp}\,dp$$

伯努力模型的极大似然估计：

$$L(p) = \binom{n}{k} p^k (1-p)^{n-k}$$

求导⇒极大似然：

$$0 = \binom{n}{k}\left[kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{(n-k-1)}\right]$$

$$\Rightarrow kp^{k-1}(1-p)^{n-k} = (n-k)p^k(1-p)^{n-k-1}$$

$$\Rightarrow k(1-p) = (n-k)p \Rightarrow p = \frac{k}{n}$$ （$p=0$或1舍去）

伯努力模型的贝叶斯估计：

$$E_\pi[p|A_1,\dots,A_n] = \int p\,\pi(p|A_1,\dots,A_n)\,dp$$

$$= \int_0^1 \frac{p\cdot p^k(1-p)^{n-k}}{\int_0^1 p^k(1-p)^{n-k}dp}\,dp$$

$$= \frac{U(k+2,n-k+1)}{U(k+1,n-k+1)}$$ 其中U代表均匀分布的概率密度函数

$$= \frac{k+1}{n+2}$$

2. 通过经验风险最小化推导极大似然估计。证明模型是条件概率分布，当损失函数是对数损失函数时，经验风险最小化等价于极大似然估计。

假设模型的条件概率分布是 $P_\theta(Y|X)$，样本集 $D=\{(x_1,y_1),\dots(x_N,y_N)\}$

极大似然估计的似然损失函数为：

$$L(Y,P_\theta(Y|X))=-\log P_\theta(Y|X)$$

经验风险最小化公式为：

$$\min_{f\in F}\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))$$

$$\therefore \arg\min_{f\in F}\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))=\arg\min_{f\in F}\frac{1}{N}\sum_{D}[-\log P_\theta(Y|X)]$$

$$=\frac{1}{N}\arg\max_{f\in F}\log\prod_{D}P_\theta(Y|X)$$

而似然函数定义为 $L(\theta)=\prod_{D}P_\theta(Y|X)$

$$\therefore \arg\min_{f\in F}\frac{1}{N}\sum_{i=1}^{N}L(y_i,f(x_i))=\frac{1}{N}\arg\max_{f\in F}\log L(\theta)$$

反之亦然得证.