Homework 7 说明

1 习题二

Q: 推导朴素贝叶斯分类方法中先验概率和条件概率的极大似然估计和贝叶斯估计。

1.1 先验概率的极大似然估计

记 $P(Y=c_k)$ 为 p_k，$\sum_{i=1}^{N} I(y_i=c_k)$ 为 n_k，对数似然函数（系数可忽略）为

$$L(p_k)=\sum_{k=1}^{K} n_k log(p_k)$$

约束条件为 $\sum_{k=1}^{K} p_k=1$，故 Lagrange 函数为

$$L(p_k,\lambda)=\sum_{k=1}^{K} n_k log(p_k)-\lambda(\sum_{k=1}^{K} p_k-1)$$

最优性条件为

$$\frac{\partial L(p_k,\lambda)}{\partial p_k}=\frac{n_k}{p_k}-\lambda=0$$

$$\frac{\partial L(p_k,\lambda)}{\partial \lambda}=1-\sum_{k=1}^{K} p_k=0$$

解得

$$\lambda=N,\ \ \hat{p_k}=\frac{n_k}{N}$$

条件概率的极大似然估计类似可得。

1.2 先验概率的贝叶斯估计

由于似然函数为多项分布，考虑多项分布的共轭先验 Dirichlet 分布。假设 $(p_1,\cdots,p_K)$ 的先验分布为以 $\alpha=(\alpha_1,\cdots,\alpha_K)$ 为参数的 Dirichlet 分布

$$Dir(p_k|\alpha)=\frac{1}{B(\alpha)}\Pi_{k=1}^{K} p_k^{\alpha_k-1}$$

其中 $B(\alpha)$ 是与 p_k 无关的常数。故 $(p_1,\cdots,p_K)$ 的后验分布满足

$$P(p_1,\cdots,p_K)\propto exp(L(p_k))\cdot Dir(p_k|\alpha)\propto \Pi_{k=1}^{K} p_k^{n_k+\alpha_k-1}$$

所以 $(p_1,\cdots,p_K)$ 的后验分布是参数为 $(n_1+\alpha_1,\cdots,n_K+\alpha_K)$ 的 Dirichlet 分布。由 Dirichlet 分布的性质，

$$E(p_k)=\frac{n_k+\alpha_k}{\sum_{k=1}^{K}(n_k+\alpha_k)}$$

取 $\alpha_1=\cdots=\alpha_K=\lambda$，即得

$$E(p_k)=\frac{n_k+\lambda}{N+K\lambda}$$

条件概率的贝叶斯估计类似可得。