朴素贝叶斯

Peng Li

https://simplelp.github.io/

2019/06/03

朴素贝叶斯分类器

$$y = f(x) = \arg\max_{c_k} \frac{P(Y = c_k)\prod_j P(X^{(j)} = x^{(j)} \mid Y = c_k)}{\sum_k P(Y = c_k)\prod_j P(X^{(j)} = x^{(j)} \mid Y = c_k)}$$

$$y = \arg\max_{c_k} P(Y = c_k)\prod_j P(X^{(j)} = x^{(j)} \mid Y = c_k)$$

参数估计： 最大似然估计

$$P(Y=c_k)=\frac{\sum_{i=1}^{N} I(y_i=c_k)}{N}, \quad k=1,2,\cdots,K$$

$$P(X^{(j)}=a_{jl} \mid Y=c_k)=\frac{\sum_{i=1}^{N} I(x_i^{(j)}=a_{jl}, y_i=c_k)}{\sum_{i=1}^{N} I(y_i=c_k)}$$

$$j=1,2,\cdots,n; \quad l=1,2,\cdots,S_j; \quad k=1,2,\cdots,K$$

参数估计：贝叶斯估计

$$P_\lambda(Y=c_k)=\frac{\sum_{i=1}^{N} I(y_i=c_k)+\lambda}{N+K\lambda}$$

$$P_\lambda(X^{(j)}=a_{jl}\mid Y=c_k)=\frac{\sum_{i=1}^{N} I(x_i^{(j)}=a_{jl},y_i=c_k)+\lambda}{\sum_{i=1}^{N} I(y_i=c_k)+S_j\lambda}$$

思考

- 证明后验概率最大化等价于期望风险最小化（可参考课本）
- 两种参数估计的证明
- 利用两种参数估计方法对例题4.1进行分类
- 代码实现朴素贝叶斯分类器

Thanks !