1.朴素贝叶斯分类

(1) 模型假设:

输入: p维特征向量

输出: 类别标记 $y \in \{c_1, c_2, \cdots, c_K\}$, 当$K=2$时, 对应二分类问题

$$T = \{(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)\}$$

(x_i, y_i) 由 $P(X, Y)$ 独立同分布产生,

目标: 学习 $P(X, Y)$

- 先验概率分布 $P(Y=c_k)$ $(k=1,2,\cdots,K)$

  条件概率 $P(X=x|Y=c_k)$

(2) 后验概率最大化

$$P(Y=c_k|X=x) = \frac{P(X=x|Y=c_k)P(Y=c_k)}{\sum_k P(X=x|Y=c_k)\ (Y=c_k)}$$

(3) 条件独立性假设

$P(X=x|Y=c_k)$ 设$x^{(j)}$是离散值, 可能取值有S_j个, Y的可能取值有K个

$$P(X=x|Y=c_k) = P(X^{(1)}=x^{(1)}, X^{(2)}=x^{(2)}, \cdots, X^{(p)}=x^{(p)}|Y=c_k)$$
$$= \prod_{j=1}^{p} P(X^{(j)}=x^{(j)}|Y=c_k)$$

这一假设使朴素贝叶斯变得简单, 但也会损失分类准确性

$$P(Y=c_k|X=x) = \frac{P(Y=c_k)\prod_{j=1}^{p} P(X^{(j)}=x^{(j)}|Y=c_k)}{\sum_{k=1}^{K} P(Y=c_k)\prod_{j=1}^{p} P(X^{(j)}=x^{(j)}|Y=c_k)}$$

由于分母都相同,

$$y = \arg\max_{y} P(Y=c_k)\prod_{j=1}^{p} P(X^{(j)}=x^{(j)}|Y=c_k)$$

注:后验概率最大化等价于期望风险最小化

(选择0-1损失) $L(Y,f(X))=\begin{cases}1 & Y\neq f(X)\\ 0 & Y=f(X)\end{cases}$

$$R_{emp}(f)=E\{L(Y,f(X))\}=E_X\{E_{Y|X}(L(Y,f(X))|X)\}$$

$$=E_X\{\sum_k L(c_k,f(x))P(Y=c_k|X)\}$$

只需对每个$X=x$每个极小化:

$$f(x)=\arg\min_y \sum_k L(c_k,y)P(c_k|X=x)$$

$$=\arg\min_y \sum_k I(c_k\neq y)P(c_k|X=x)$$

$$=\arg\min_y P(Y\neq y|X=x)=\arg\max_y P(Y=y|X=x)$$

2. 参数估计

2.1 极大似然估计

$$P(Y=c_k)=\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N}$$

每个特征可能的取值集合 $\{a_{j1},a_{j2},\cdots,a_{jS_j}\}$

$$P(X^{(j)}=a_{jl}|Y=c_k)=\frac{\sum_i I(x_i^{(j)}=a_{jl},y_j=c_k)}{\sum_i I(y_j=c_k)}$$

2.2 贝叶斯估计

$$P_\lambda(X^{(j)}=a_{jl}|Y=c_k)=\frac{\sum_i I(x_i^{(j)}=a_{jl},y_j=c_k)+\lambda}{\sum_i I(y_j=c_k)+S_j\cdot\lambda}$$

$$P_\lambda(Y=c_k)=\frac{\sum_i I(y_i=c_k)+\lambda}{N+K\lambda}$$