统计学习方法

第四章　朴素贝叶斯法（Naive Bayes）

黄嘉锋

2018.10.26

第四章 朴素贝叶斯法

4.1.1 基本方法

- 定义 ：
  - 输入空间：$\mathcal{X} \subseteq R^n$，输出空间：$\mathcal{Y} = \{c_1, c_2, \dots, c_k\}$
  - 输入输出：特征向量$x \in \mathcal{X}$，类标记$y \in \mathcal{Y}$
  - X是定义在$\mathcal{X}$上的随机向量
  - Y是定义在$\mathcal{Y}$上的随机向量
  - 训练数据集：$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

4.1.1 基本方法

- 概率定义 ：
  - 先验概率： $P(Y = c_k), k = 1,2, \dots, K$
  - 条件概率：

$$P(X = x|Y = c_k) = P\left(X^{(1)} = x^{(1)}, \dots, X^{(n)} = x^{(n)} \middle| Y = c_k\right),$$
$$k = 1,2, \dots, K$$

  - 联合概率： $P(X, Y)$

4.1.1 基本方法

- 条件概率分布$P(X=x|Y=c_k)$有指数数量的参数：
  - 假设$x^{(j)}$可取值有S_j个， $j=1,2,\dots,n$， Y可取值有K个，那么参数个数为

$$K\prod_{j=1}^{n}S_j$$

4.1.1 基本方法

- 条件独立性假设→

$$P(X=x|Y=c_k)= P\left(X^{(1)}=x^{(1)},\ldots,X^{(n)}=x^{(n)}\middle|Y=c_k\right)$$
$$=\prod_{j=1}^{n}P(X^{(j)}=x^{(j)}|Y=c_k)$$

4.1.1 基本方法

- 将后验概率$P(Y=c_k|X=x)$最大的类作为x的类输出：

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

$$=\frac{P(Y=c_k)\prod_j P(X^{(j)}=x^{(j)}|Y=c_k)}{\sum_k P(Y=c_k)\prod_j P(X^{(j)}=x^{(j)}|Y=c_k)}, \quad k=1,2,\ldots,K$$

- 其中分母对所有c_k都是相同的

4.1.1 基本方法

- 朴素贝叶斯分类器简化为：

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

4.1.2 后验概率最大化的含义

- 0-1损失函数： $L\big(Y,f(X)\big)=\begin{cases}1, Y\neq f(X)\\0, Y=f(X)\end{cases}$

- 期望风险函数： $R_{exp}(f)=E[L(Y,f(X))]$

- 条件期望： $R_{exp}(f)=E_x\sum_{k=1}^{K}[L(c_k,f(X))]P(c_k|X)$

4.1.2 后验概率最大化的含义

- 使期望风险最小化，对$X = x$逐个极小化：

$$f(x) = arg \min_{y \in \mathcal{Y}} \sum_{k=1}^{K} L(c_k, y) P(c_k | X = x)$$

$$= arg \min_{y \in \mathcal{Y}} \sum_{k=1}^{K} P(y \neq c_k | X = x)$$

$$= arg \min_{y \in \mathcal{Y}} (1 - P(y = c_k | X = x))$$

$$= arg \max_{y \in \mathcal{Y}} P(y = c_k | X = x)$$

第四章 朴素贝叶斯法

4.2.1 极大似然估计

- 先验概率$P(Y=c_k)$:

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

- 设$x^{(j)}$可能取值为$\{a_{j1},a_{j2},\dots,a_{jS_j}\}$, 条件概率$P\left(X^{(j)}=a_{jl}\middle|Y=c_k\right)$:

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

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

4.2.2 学习与分类算法

- 计算先验概率及条件概率：

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

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

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

4.2.2 学习与分类算法

- 对于给定的实例$x = (x^{(1)},\ x^{(2)}, \dots, x^{(n)})^T$, 计算：

$$P(Y = c_k)\prod_j P(X^{(j)} = x^{(j)}|Y = c_k),\quad k = 1,2,\dots,K$$

- 确定x的类：

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

4.2.3 贝叶斯估计

- 先验概率：

$$P_\lambda(Y=c_k)=\frac{\sum_{i=1}^{N} I(y_i=c_k)+\lambda}{N+K\lambda}, k=1,2,\ldots,K$$

- 条件概率：

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

- $\lambda \geq 0, \lambda=0$为极大似然估计，$\lambda=1$为拉普拉斯平滑(Laplace smoothing)

第四章 朴素贝叶斯法

4.1 朴素贝叶斯法的学习与分类

4.2 朴素贝叶斯法的参数估计

4.3 总结

4.3 总结

- 朴素贝叶斯法的生成方法
- 朴素贝叶斯法的基本假设
- 期望角度的朴素贝叶斯法