第四章

朴素贝叶斯法

张志忠 副教授

华东师范大学计算机学院

目录

一、朴素贝叶斯法的学习与分类

- 基本方法
- 后验概率最大化的含义

基本方法

- 训练数据集：

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$$

- 由X和Y的联合概率分布P(X,Y)独立同分布产生

- 朴素贝叶斯通过训练数据集学习联合概率分布P(X,Y)，
  - 即先验概率分布：$P(Y=c_k)$， $k=1,2,\cdots,K$
  - 及条件概率分布：

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

  - 注意：条件概率为指数级别的参数：

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

基本方法

- 条件独立性假设：

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

- “朴素”贝叶斯名字由来，牺牲分类准确性。

- 贝叶斯定理：$P(Y=c_k|X=x)=\dfrac{P(X=x|Y=c_k)P(Y=c_k)}{\sum_k P(X=x|Y=c_k)P(Y=c_k)}$
- 代入上式：

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

基本方法

- 贝叶斯分类器：

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

- 分母对所有c_k都相同：

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

后验概率最大化的含义：

- 朴素贝叶斯法将实例分到后验概率最大的类中，等价于期望风险最小化，
- 假设选择0-1损失函数：f(X)为决策函数

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

- 期望风险函数：

$$R_{\exp}(f)=E[L(Y,f(X))]$$

- 取条件期望：

$$R_{\exp}(f)=E_X\sum_{k=1}^{K}[L(c_k,f(X))]P(c_k\mid X)$$

后验概率最大化的含义：

- 只需对X=x逐个极小化，得：

$$
\begin{aligned}
f(x) &= \arg\min_{y \in \mathcal{Y}} \sum_{k=1}^{K} L(c_k, y) P(c_k \mid X = x) \\
&= \arg\min_{y \in \mathcal{Y}} \sum_{k=1}^{K} P(y \neq c_k \mid X = x) \\
&= \arg\min_{y \in \mathcal{Y}} (1 - P(y = c_k \mid X = x)) \\
&= \arg\max_{y \in \mathcal{Y}} P(y = c_k \mid X = x)
\end{aligned}
$$

- 推导出后验概率最大化准则：

$$
f(x) = \arg\max_{c_k} P(c_k \mid X = x)
$$

二、朴素贝叶斯法的参数估计

- 应用极大似然估计法估计相应的概率：
- 先验概率$P(Y=c_k)$的极大似然估计是：

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

- 设第j个特征$x^{(j)}$可能取值的集合为：$\{a_{j1},a_{j2},\cdots,a_{jS_j}\}$
- 条件概率的极大似然估计：

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

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

朴素贝叶斯法的参数估计

- 学习与分类算法Naïve Bayes Algorithm：
- 输入：
  - 训练数据集 $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$
  - $x_i^{(j)}$ 第i个样本的第j个特征 $x_i=(x_i^{(1)},x_i^{(2)},\cdots,x_i^{(n)})^{\mathrm{T}}$
  - a_{jl} 第j个特征可能取的第l个值 $x_i^{(j)}\in\{a_{j1},a_{j2},\cdots,a_{jS_j}\}$

$$y_i\in\{c_1,c_2,\cdots,c_K\}$$

- 输出：
  - x的分类

朴素贝叶斯法的参数估计

- 步骤
  - 1、计算先验概率和条件概率

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

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

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

朴素贝叶斯法的参数估计

- 步骤
  - 2、对于给定的实例 $x=(x^{(1)},x^{(2)},\cdots,x^{(n)})^{\mathrm{T}}$
  - 计算

$$P(Y=c_k)\prod_{j=1}^{n}P(X^{(j)}=x^{(j)}\mid Y=c_k),\quad k=1,2,\cdots,K$$

  - 3、确定x的类别

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

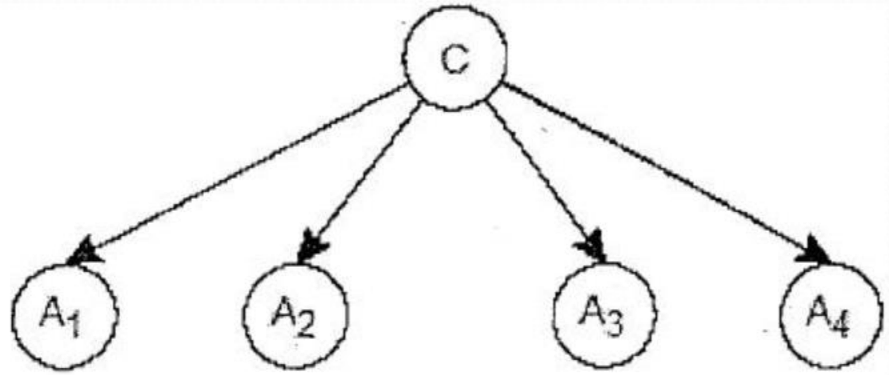
例子

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

例子

- 测试

<*Outlook=sunny, Temperature=cool, Humidity=high, Wind=strong*>



$$c(x) = \arg\max_{c\in\{yes,no\}} P(c)P(sunny \mid c)P(cool \mid c)P(high \mid c)P(strong \mid c)$$

例子

$P(yes)=(9+1)/(14+2)=10/16$

$P(sunny \mid yes)=(2+1)/(9+3)=3/12$

$P(cool \mid yes)=(3+1)/(9+3)=4/12$

$P(high \mid yes)=(3+1)/(9+2)=4/11$

$P(strong \mid yes)= (3+1)/(9+2)=4/11$

$P(no)=(5+1)/(14+2)=6/16$

$P(sunny \mid no)=(3+1)/(5+3)=4/8$

$P(cool \mid no)=(1+1)/(5+3)=2/8$

$P(high \mid no)=(4+1)/(5+2)=5/7$

$P(strong \mid no)=(3+1)/(5+2)=4/7$

$$P(yes)P(sunny|yes)P(cool|yes)P(high|yes)P(strong|yes)=0.0069$$

$$P(no)P(sunny|no)P(cool|no)P(high|no)P(strong|no)=0.0191$$

贝叶斯估计

- 考虑：用极大似然估计可能会出现所要估计的概率值为0的情况，这时会影响到后 验概率的计算结果，使分类产生偏差.解决这一问题的方法是采用贝叶斯估计。
- 条件概率的贝叶斯估计：

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

- 先验概率的贝叶斯估计：

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

Q & A