

华中科技大学

机器学习导论作业报告

专业院系 : 电子信息与通信学院
专业班级 : 电信1805班
学生学号 : U201813372
学生姓名 : 朱浩然

第四章 贝叶斯分类器

1. 简述朴素贝叶斯的优缺点.

优点:

- (1) 朴素贝叶斯模型发源于古典数学理论，有稳定的分类效率。
- (2) 对小规模的数据表现很好，能个处理多分类任务，适合增量式训练，尤其是数据量超出内存时，我们可以一批批的去增量训练。
- (3) 对缺失数据不太敏感，算法也比较简单，常用于文本分类。

缺点:

- (1) 理论上，朴素贝叶斯模型与其他分类方法相比具有最小的误差率。但是实际上并非总是如此，这是因为朴素贝叶斯模型给定输出类别的情况下，假设属性之间相互独立，这个假设在实际应用中往往是不成立的，在属性个数比较多或者属性之间相关性较大时，分类效果不好。而在属性相关性较小时，朴素贝叶斯性能最为良好。对于这一点，有半朴素贝叶斯之类的算法通过考虑部分关联性适度改进。
- (2) 需要知道先验概率，且先验概率很多时候取决于假设，假设的模型可以有很多种，因此在某些时候会由于假设的先验模型的原因导致预测效果不佳。
- (3) 由于我们是通过先验和数据来决定后验的概率从而决定分类，所以分类决策存在一定的错误率。
- (4) 对输入数据的表达形式很敏感。

2. 试由下表的训练数据学习一个朴素贝叶斯分类器并确定 $x = (2, S)^T$ 的类标记 y . 表中 $X^{(1)}$, $X^{(2)}$ 为特征, 取值的集合分别为 $A_1 = \{1, 2, 3\}$, $A_2 = \{S, M, L\}$, Y 为类标记, $Y \in C = \{1, -1\}$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

$$P(Y=1) = \frac{9}{15}, P(Y=-1) = \frac{6}{15}$$

$$P(X^{(1)}=1|Y=1) = \frac{2}{9}, P(X^{(1)}=2|Y=1) = \frac{3}{9}, P(X^{(1)}=3|Y=1) = \frac{4}{9}$$

$$P(X^{(2)}=S|Y=1) = \frac{1}{9}, P(X^{(2)}=M|Y=1) = \frac{4}{9}, P(X^{(2)}=L|Y=1) = \frac{4}{9}$$

$$P(X^{(1)}=1|Y=-1) = \frac{3}{6}, P(X^{(1)}=2|Y=-1) = \frac{2}{6}, P(X^{(1)}=3|Y=-1) = \frac{1}{6}$$

$$P(X^{(2)}=S|Y=-1) = \frac{3}{6}, P(X^{(2)}=M|Y=-1) = \frac{2}{6}, P(X^{(2)}=L|Y=-1) = \frac{1}{6}$$

对于给定的 $x = (2, S)^T$ 计算:

$$P(Y=1)P(X^{(1)}=2|Y=1)P(X^{(2)}=S|Y=1) = \frac{9}{15} * \frac{3}{9} * \frac{1}{9} = \frac{1}{45}$$

$$P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1) = \frac{6}{15} * \frac{2}{6} * \frac{3}{6} = \frac{1}{15}$$

因为 $P(Y=-1)P(X^{(1)}=2|Y=-1)P(X^{(2)}=S|Y=-1)$ 最大，所以 $y=-1$

3. 考虑下表中的数据

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	-
3	0	1	1	-
4	0	1	1	-
5	0	0	1	+
6	1	0	1	+
7	1	0	1	-
8	1	0	1	-
9	1	1	1	+
10	1	0	1	+

(a) 估计条件概率 $P(A|+)$, $P(B|+)$, $P(C|+)$, $P(A|-)$, $P(B|-)$ 和 $P(C|-)$.

$$\textcircled{1} P(A=1 | +) = \frac{3}{5} \quad \textcircled{2} P(B=1 | +) = \frac{1}{5} \quad \textcircled{3} P(C=1 | +) = \frac{4}{5}$$

$$\textcircled{4} P(A=1 | -) = \frac{2}{5} \quad \textcircled{5} P(B=1 | -) = \frac{2}{5} \quad \textcircled{6} P(C=1 | -) = 1$$

$$\textcircled{7} P(A=0 | +) = \frac{2}{5} \quad \textcircled{8} P(B=0 | +) = \frac{4}{5} \quad \textcircled{9} P(C=0 | +) = \frac{1}{5}$$

$$\textcircled{10} P(A=0 | -) = \frac{3}{5} \quad \textcircled{11} P(B=0 | -) = \frac{3}{5} \quad \textcircled{12} P(C=0 | -) = 0$$

(b) 根据(a)中的条件概率，使用朴素贝叶斯方法预测测试样本(A=0, B=1, C=0)的类标号

假设 $P(A=0, B=1, C=0) = K$ ，则K属于两个类的概率为：

$$\text{贝叶斯公式 } P(\text{类别}|\text{特征}) = \frac{P(\text{特征}|\text{类别}) * P(\text{类别})}{P(\text{特征})}$$

$$\begin{aligned} \textcircled{1} P(+ | A=0, B=1, C=0) &= P(A=0, B=1, C=0) * P(+)/K \\ &= P(A=0 | +)P(B=1 | +)P(C=0 | +) \times P(+)/K \\ &= 0.4 * 0.2 * 0.2 * 0.5 / K = 0.008 / K \end{aligned}$$

$$\begin{aligned} \textcircled{2} P(- | A=0, B=1, C=0) &= P(A=0, B=1, C=0) * P(-)/K \\ &= P(A=0 | -)P(B=1 | -)P(C=0 | -) \times P(-)/K \\ &= 0.4 * 0.2 * 0 * 0.5 / K = 0 / K \end{aligned}$$

因此可以得到，此样本的类标号是：+

(c) 使用m估计方法(p=1/2 且 m=4)估计条件概率

$$\text{M-估计公式: } P = (r + P_a * m) / (n + m)$$

$$\textcircled{1} P(A=1 | +) = \frac{3+2}{5+4} = \frac{5}{9} \quad \textcircled{2} P(B=1 | +) = \frac{1+2}{5+4} = \frac{1}{3} \quad \textcircled{3} P(C=1 | +) = \frac{4+2}{5+4} = \frac{2}{3}$$

$$\textcircled{4} P(A=1 | -) = \frac{2+2}{5+4} = \frac{4}{9} \quad \textcircled{5} P(B=1 | -) = \frac{2+2}{5+4} = \frac{4}{9} \quad \textcircled{6} P(C=1 | -) = \frac{5+2}{5+4} = \frac{7}{9}$$

$$\textcircled{7} P(A=0 \mid +) = \frac{4}{9} \textcircled{8} P(B=0 \mid +) = \frac{2}{3} \textcircled{9} P(C=0 \mid +) = \frac{1}{3}$$

$$\textcircled{10} P(A=0 \mid -) = \frac{5}{9} \textcircled{11} P(B=0 \mid -) = \frac{5}{9} \textcircled{12} P(C=0 \mid -) = \frac{2}{9}$$

(d) 同(b)，使用(c)中的条件概率

假设 $P(A=0, B=1, C=0) = K$ ，则 K 属于两个类的概率为：

$$\begin{aligned} \textcircled{1} P(+ \mid A=0, B=1, C=0) &= P(A=0, B=1, C=0) \cdot P(+)/K \\ &= P(A=0 \mid +) P(B \mid +) P(C=0 \mid +) \times P(+)/K \\ &= \frac{4}{9} \times \frac{1}{3} \times \frac{1}{3} \times 0.5/K = 0.0247/K \end{aligned}$$

$$\begin{aligned} \textcircled{2} P(- \mid A=0, B=1, C=0) &= P(A=0, B=1, C=0) \cdot P(-)/K \\ &= P(A=0 \mid -) P(B \mid -) P(C=0 \mid -) \times P(-)/K \\ &= \frac{5}{9} \times \frac{4}{9} \times \frac{2}{9} \times 0.5/K = 0.0274/K \end{aligned}$$

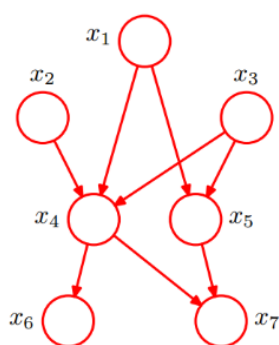
因此可以得到，此样本的类标号是：-

(e) 比较估计概率的两种方法。哪一种更好？为什么？

当条件概率有为 0 的时候，用 M-估计的方法比条件概率的贝叶斯更好，因为我们不想整个概率计算结果为 0，这样一般是不符合实际的。M-估计可以避免概率计算结果为 0 的情况

4. 给定如下图所示的一个贝叶斯网络

- 请写出 x_1, x_2, \dots, x_7 的联合概率分布
- x_1 和 x_2 是否相互独立？
- x_6 和 x_7 在给定 x_4 的条件下是否相互独立？



(a) 由上图可以看出 $x_1, x_2, x_3, \dots, x_7$ 的联合概率分布为：

$$P(x_1) \cdot P(x_2) \cdot P(x_3) \cdot P(x_4 \mid x_1, x_2, x_3) \cdot P(x_5 \mid x_1, x_3) \cdot P(x_6 \mid x_4) \cdot P(x_7 \mid x_4, x_5)$$

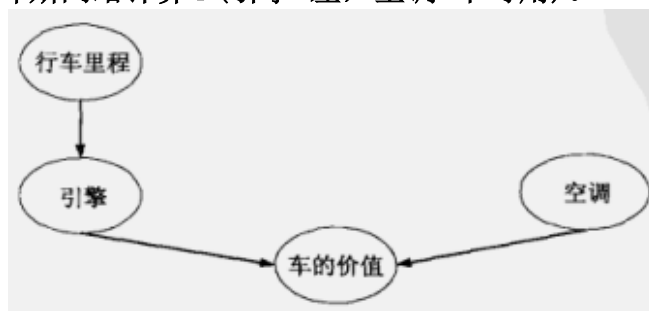
(b) x_1 和 x_2 独立 (对应 head-to-head)，因为 x_1 和 x_2 都是两个 head，head-head 之间是相互独立的

(c) x_6 和 x_7 在 x_4 给定的条件下独立 (对应 tail-to-tail)；因为 x_6 的概率只与 x_4 的概率有关， x_7 的概率与 x_4, x_5 有关，若给定 x_4 的条件下，显

然此时 x_7 的概率取决于 x_5 ，与 x_6 无关， x_6-x_7 等价于 tail-tail，二者相互独立

5. 下图给出了表中的数据集对应的贝叶斯信念网络（假设所有的属性都是二元的）。

- (a) 画出网络中每个结点对应的概率表。
(b) 使用贝叶斯网络计算 $P(\text{引擎}=\text{差}, \text{空调}=\text{不可用})$ 。



贝叶斯信念网络

行车里程	引擎	空调	车的价值=高的记录数	车的价值=低的记录数
高	好	可用	3	4
高	好	不可用	1	2
高	差	可用	1	5
高	差	不可用	0	4
低	好	可用	9	0
低	好	不可用	5	1
低	差	可用	1	2
低	差	不可用	0	2

数据集

① 行车里程

行车里程=高	行车里程=低
0.5	0.5

② 空调

空调=可用	空调=不可用
5/8	3/8

③ 引擎

	引擎=好	引擎=差
行车里程=高	0.5	0.5
行车里程=低	0.75	0.25

④ 车的价值

引擎	空调	车的价值=高	车的价值=低
----	----	--------	--------

好	可用	12/16=0.75	4/16=0.25
好	不可用	6/9=2/3	3/9=1/3
差	可用	2/9	7/9
差	不可用	0	1

(b) 计算过程如下:

$$P(\text{引擎} = \text{差}, \text{空调} = \text{不可用}) = P(\text{引擎} = \text{差}) * P(\text{空调} = \text{不可用})$$

$$= \{P(\text{引擎} = \text{差} | \text{行车里程} = \text{高}) \times P(\text{行车里程} = \text{高}) + \\ P(\text{引擎} = \text{差} | \text{行车里程} = \text{低}) \times P(\text{行车里程} = \text{低})\} \\ \times P(\text{空调} = \text{不可用})$$

$$= (0.5 * 0.5 + 0.25 * 0.5) * 0.375 = 0.140625$$