

# 机器学习第七章作业

---

161910126 赵安

## 题目1

(1) 通过查表得,  $x = \{0, B\}$  的类别是 “0”

(2)

首先估计类先验概率  $P(c)$ , 显然有

$$P(y = 0) = \frac{6}{15} = 0.4$$

$$P(y = 1) = \frac{9}{15} = 0.6$$

然后, 为每个属性估计条件概率  $P(x_i|c)$  :

$$P(0|0) = \frac{2}{6} = 0.333$$

$$P(0|1) = \frac{3}{9} = 0.333$$

$$P(B|0) = \frac{3}{6} = 0.500$$

$$P(B|1) = \frac{1}{9} = 0.111$$

于是, 有

$$P(y = 1) \times P_{0|1} \times P_{B|1} \approx 0.022$$

$$P(y = 0) \times P_{0|0} \times P_{B|0} \approx 0.067$$

由于  $0.067 > 0.022$ , 因此, 朴素贝叶斯分类器将测试样本分类判别为 “0”。

(3)

使用“拉普拉斯修正”

$$P(y=0) = \frac{6+1}{15+2} \approx 0.412$$

$$P(y=1) = \frac{9+1}{15+2} \approx 0.588$$

然后，为每个属性估计条件概率  $P(x_i|c)$ ：

$$P(0|0) = \frac{2+1}{6+3} = 0.333$$

$$P(0|1) = \frac{3+1}{9+3} = 0.333$$

$$P(B|0) = \frac{3+1}{6+3} = 0.444$$

$$P(B|1) = \frac{1+1}{9+3} = 0.167$$

于是，有

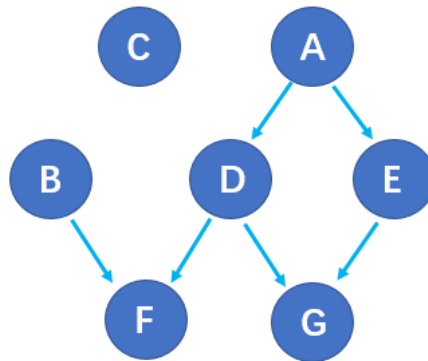
$$P(y=1) \times P_{0|1} \times P_{B|1} \approx 0.033$$

$$P(y=0) \times P_{0|0} \times P_{B|0} \approx 0.061$$

由于  $0.061 > 0.033$ ，因此，朴素贝叶斯分类器将测试样本分类判别为“0”。

## 题目2

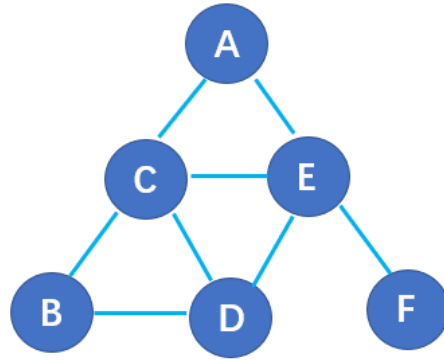
(1) 贝叶斯网结构如下：



(2)

$$Pr(A, B, C, D, E, F) = Pr(A)Pr(B)Pr(C|A, B)Pr(D|B)Pr(E|C, D)Pr(F|E)$$

(3) 道德图如下：



序号	关系	True/False	序号	关系	True/False
1	$A \not\perp B$	T	7	$F \perp B   C$	F
2	$A \perp B   C$	F	8	$F \perp B   C, D$	T
3	$C \not\perp D$	T	9	$F \perp B   E$	T
4	$C \perp D   E$	F	10	$A \not\perp F$	F
5	$C \perp D   B, F$	F	11	$A \perp F   C$	F
6	$F a \not\perp B$	F	12	$A \perp F   D$	F

备注：\upmodels 符号显示不出来

### 题目3

试编程实现拉普拉斯修正的朴素贝叶斯分类器，并以西瓜数据集 3.0 为训练集，对 p.151 "测 1" 样本进行判别。

```

import math

# 定义数据集
dataset = [
    ['青绿', '蜷缩', '浊响', '清晰', '凹陷', '硬滑', 0.697, 0.460, 1],
    ['乌黑', '蜷缩', '沉闷', '清晰', '凹陷', '硬滑', 0.774, 0.376, 1],
    ['乌黑', '蜷缩', '浊响', '清晰', '凹陷', '硬滑', 0.634, 0.264, 1],
    ['青绿', '蜷缩', '沉闷', '清晰', '凹陷', '硬滑', 0.608, 0.318, 1],
    ['浅白', '蜷缩', '浊响', '清晰', '凹陷', '硬滑', 0.556, 0.215, 1],
    ['青绿', '稍蜷', '浊响', '清晰', '稍凹', '软粘', 0.403, 0.237, 1],
    ['乌黑', '稍蜷', '浊响', '稍糊', '稍凹', '软粘', 0.481, 0.149, 1],
    ['乌黑', '稍蜷', '浊响', '清晰', '稍凹', '硬滑', 0.437, 0.211, 1],
    ['乌黑', '稍蜷', '沉闷', '稍糊', '稍凹', '硬滑', 0.666, 0.091, 0],
    ['青绿', '硬挺', '清脆', '清晰', '平坦', '软粘', 0.243, 0.267, 0],
    ['浅白', '硬挺', '清脆', '模糊', '平坦', '硬滑', 0.245, 0.057, 0],
    ['浅白', '蜷缩', '浊响', '模糊', '平坦', '软粘', 0.343, 0.099, 0],
    ['青绿', '稍蜷', '浊响', '稍糊', '凹陷', '硬滑', 0.639, 0.161, 0],
    ['浅白', '稍蜷', '沉闷', '稍糊', '凹陷', '硬滑', 0.657, 0.198, 0],
    ['乌黑', '稍蜷', '浊响', '清晰', '稍凹', '软粘', 0.360, 0.370, 0],
    ['浅白', '蜷缩', '浊响', '模糊', '平坦', '硬滑', 0.593, 0.042, 0],
    ['青绿', '蜷缩', '沉闷', '稍糊', '稍凹', '硬滑', 0.719, 0.103, 0]
]

# 求每种特征的属性个数

```

```

uiq = []
temp = []
for j in range(6):
    for i in range(len(dataSet)):
        temp.append(dataSet[i][j])
        temp = list(set(temp))
    uiq.append(len(temp))

# 对连续属性考虑概率密度函数
def density_1(x):
    return 3.092624 * math.exp(-((x - 0.574) ** 2) / 0.033282))

def density_0(x):
    return 2.045857 * math.exp(-((x - 0.496) ** 2) / 0.07605))

def sugar_1(x):
    return 3.949923 * math.exp(-((x - 0.279) ** 2) / 0.020402))

def sugar_0(x):
    return 3.693910 * math.exp(-((x - 0.154) ** 2) / 0.023328))

# 带拉普拉斯修正的朴素贝叶斯分类器
def naive_bayes(test):
    # 将 好瓜 与 坏瓜 分开
    GW = []
    BW = []
    for i in range(len(dataSet)):
        if dataSet[i][8] == 1:
            GW.append(dataSet[i])
        else:
            BW.append(dataSet[i])
    # 初始化好瓜与坏瓜的概率
    pg = 1.0
    pb = 1.0

    # 估计类先验概率
    P_D1 = len(GW) / len(dataSet)
    P_D0 = len(BW) / len(dataSet)

    # 为每个属性估计条件概率
    for j in range(6):
        n1 = 0.0
        n0 = 0.0
        for k in range(len(GW)):
            if GW[k][j] == test[j]:
                n1 = n1 + 1.0
        pg = pg * ((n1 + 1.0) / (len(GW) + uiq[j]))

        for k in range(len(BW)):
            if BW[k][j] == test[j]:
                n0 = n0 + 1.0
        pb = pb * ((n0 + 1.0) / (len(BW) + uiq[j]))

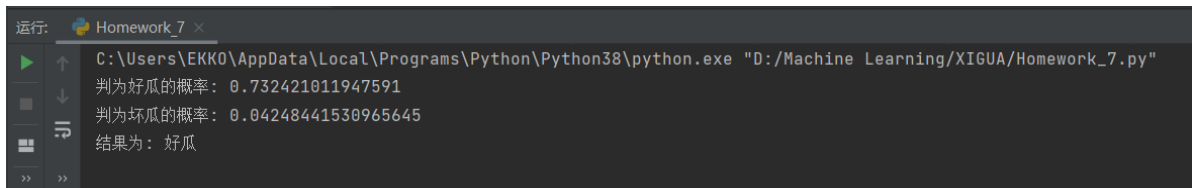
```

```
pg = P_D1 * density_1(test[6]) * sugar_1(test[7])
pb = P_D0 * density_0(test[6]) * sugar_0(test[7])

print("判为好瓜的概率:", pg)
print("判为坏瓜的概率:", pb)
result = "好瓜" if pg > pb else "坏瓜"
print("结果为:", result)

if __name__ == '__main__':
    test = ['青绿', '蜷缩', '浊响', '清晰', '凹陷', '硬滑', 0.697, 0.460]
    naive_bayes(test)
```

输出结果为：



```
运行: Homework_7 x
C:\Users\EKK0\AppData\Local\Programs\Python\Python38\python.exe "D:/Machine Learning/XIGUA/Homework_7.py"
判为好瓜的概率: 0.732421011947591
判为坏瓜的概率: 0.04248441530965645
结果为: 好瓜
```