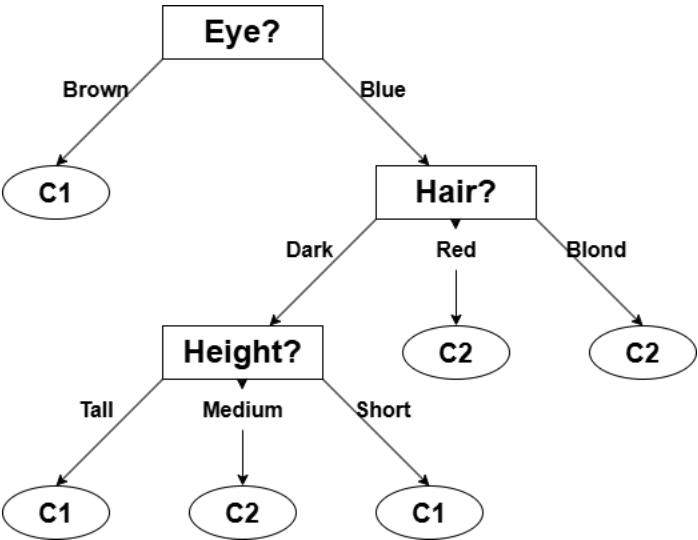


# 第二次作业

姓名：王泽黎  
学号：2022K8009929011

## 1

1.  $I(5, 4) \approx 0.991$   
 $E(\text{Height}) \approx 0.846$   
 $E(\text{Hair}) \approx 0.805$   
 $E(\text{Eye}) \approx 0.612$   
 $\text{Gain}(\text{Height}) = I(5, 4) - E(\text{Height}) = 0.145$   
 $\text{Gain}(\text{Hair}) = I(5, 4) - E(\text{Hair}) = 0.186$   
 $\text{Gain}(\text{Eye}) = I(5, 4) - E(\text{Eye}) = 0.379$
2. Decision tree constructed with Information Gain:



## 2

- 输入层设计:
  - 身高 (Height):
    - 高 (Tall): [1, 0, 0]
    - 中 (Medium): [0, 1, 0]
    - 矮 (Short): [0, 0, 1]
  - 头发 (Hair):
    - 金色 (Blond): [1, 0, 0]
    - 黑色 (Dark): [0, 1, 0]
    - 红色 (Red): [0, 0, 1]
  - 眼睛 (Eye):
    - 棕色 (Brown): [1, 0]
    - 蓝色 (Blue): [0, 1]
  - 共有 8 个输入节点
    - $I_1, I_2, I_3$  表示身高
    - $I_4, I_5, I_6$  表示头发
    - $I_7, I_8$  表示眼睛
- 输出层设计:
  - 班级 (Class):
    - 班级 1: [1, 0]
    - 班级 2: [0, 1]

- 共有 2 个输出节点
  - $O_1$  表示班级 1
  - $O_2$  表示班级 2
- 隐藏层设计:
  - 根据经验公式法, 设计隐含层节点数为 4 个, 分别是  $H_1, H_2, H_3, H_4$
- 激活函数均选用 Sigmoid 函数
- 初始参数设置:
  - 初始权重:
    - 输入层到隐含层的权重矩阵  $W_1$ :
 
$$W_1 = \begin{bmatrix} w_{11}^1 & w_{21}^1 & w_{31}^1 & w_{41}^1 & w_{51}^1 & w_{61}^1 & w_{71}^1 & w_{81}^1 \\ w_{12}^1 & w_{22}^1 & w_{32}^1 & w_{42}^1 & w_{52}^1 & w_{62}^1 & w_{72}^1 & w_{82}^1 \\ w_{13}^1 & w_{23}^1 & w_{33}^1 & w_{43}^1 & w_{53}^1 & w_{63}^1 & w_{73}^1 & w_{83}^1 \\ w_{14}^1 & w_{24}^1 & w_{34}^1 & w_{44}^1 & w_{54}^1 & w_{64}^1 & w_{74}^1 & w_{84}^1 \end{bmatrix}$$
    - 隐含层到输出层的权重矩阵  $W_2$ :
 
$$W_2 = \begin{bmatrix} w_{11}^2 & w_{21}^2 & w_{31}^2 & w_{41}^2 \\ w_{12}^2 & w_{22}^2 & w_{32}^2 & w_{42}^2 \end{bmatrix}$$
    - 不妨设  $W_1$  和  $W_2$  的初始值均为 0.1(便于计算, 实际中可随机初始化)
  - 初始偏置:
    - 输入层到隐含层的偏置  $b_1$ :
 
$$b_1 = \begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \\ b_4^1 \end{bmatrix}$$
    - 隐含层到输出层的偏置  $b_2$ :
 
$$b_2 = \begin{bmatrix} b_1^2 \\ b_2^2 \end{bmatrix}$$
    - 不妨设  $b_1$  和  $b_2$  的初始值均为 0.1(便于计算, 实际中可随机初始化)
  - 初始学习率  $\eta = 0.1$
- 对训练样本(Medium, Blond, Blue, C1):
  - 前向传播:
    - 输入样本编码:
 
$$X = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$
    - 隐含层输入:
      - $H_{in} = W_1 \cdot X + b_1$
      - $H_{in} = \begin{bmatrix} 0.4 \\ 0.4 \\ 0.4 \\ 0.4 \end{bmatrix}$
    - 隐含层输出:
      - $H = \text{sigmod}(W_1 \cdot X + b_1)$
      - $H = \begin{bmatrix} 0.598 \\ 0.598 \\ 0.598 \\ 0.598 \end{bmatrix}$
    - 输出层输入:
      - $O_{in} = W_2 \cdot H + b_2$
      - $O_{in} = \begin{bmatrix} 0.339 \\ 0.339 \end{bmatrix}$
    - 输出层输出:
      - $O = \text{sigmod}(W_2 \cdot H + b_2)$
      - $O = \begin{bmatrix} 0.584 \\ 0.584 \end{bmatrix}$
  - 反向传播:

- 目标输出:
  - $T = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
- 输出层误差:
  - $\delta O = T - O$
  - $\delta O = \begin{bmatrix} 0.416 \\ -0.584 \end{bmatrix}$
- 隐含层误差:
  - $\delta H = W_2^T \cdot \delta O * H * (1 - H)$
  - $\delta H = \begin{bmatrix} -0.004 \\ -0.004 \\ -0.004 \\ -0.004 \end{bmatrix}$
- 权重更新:
  - $W_2 = W_2 + \eta * \delta O \cdot H^T$
  - $W_1 = W_1 + \eta * \delta H \cdot X^T$
  - $W_2 = \begin{bmatrix} 0.1 + 0.1 \cdot 0.416 \cdot 0.598 & 0.1 + 0.1 \cdot -0.584 \cdot 0.598 \\ 0.1 + 0.1 \cdot 0.416 \cdot 0.598 & 0.1 + 0.1 \cdot -0.584 \cdot 0.598 \end{bmatrix}$   
 即为  
 $W_2 = \begin{bmatrix} 0.1249 & 0.0652 \\ 0.1249 & 0.0652 \end{bmatrix}$
  - $W_1 = \begin{bmatrix} 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 \\ 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 \\ 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 \\ 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 & 0.1 + 0.1 \cdot -0.004 \cdot 0 & 0.1 + 0.1 \cdot -0.004 \cdot 1 \end{bmatrix}$   
 即为  
 $W_1 = \begin{bmatrix} 0.1 & 0.0996 & 0.1 & 0.0996 & 0.1 & 0.1 & 0.1 & 0.0996 \\ 0.1 & 0.0996 & 0.1 & 0.0996 & 0.1 & 0.1 & 0.1 & 0.0996 \\ 0.1 & 0.0996 & 0.1 & 0.0996 & 0.1 & 0.1 & 0.1 & 0.0996 \\ 0.1 & 0.0996 & 0.1 & 0.0996 & 0.1 & 0.1 & 0.1 & 0.0996 \end{bmatrix}$
- 偏置更新:
  - $b_2 = b_2 + \eta * \delta O$
  - $b_1 = b_1 + \eta * \delta H$
  - $b_2 = \begin{bmatrix} 0.1 + 0.1 \cdot 0.416 \\ 0.1 + 0.1 \cdot -0.584 \end{bmatrix}$   
 即为  
 $b_2 = \begin{bmatrix} 0.1416 \\ 0.0416 \end{bmatrix}$
  - $b_1 = \begin{bmatrix} 0.1 + 0.1 \cdot -0.004 \\ 0.1 + 0.1 \cdot -0.004 \\ 0.1 + 0.1 \cdot -0.004 \\ 0.1 + 0.1 \cdot -0.004 \end{bmatrix}$   
 即为  
 $b_1 = \begin{bmatrix} 0.0996 \\ 0.0996 \\ 0.0996 \\ 0.0996 \end{bmatrix}$

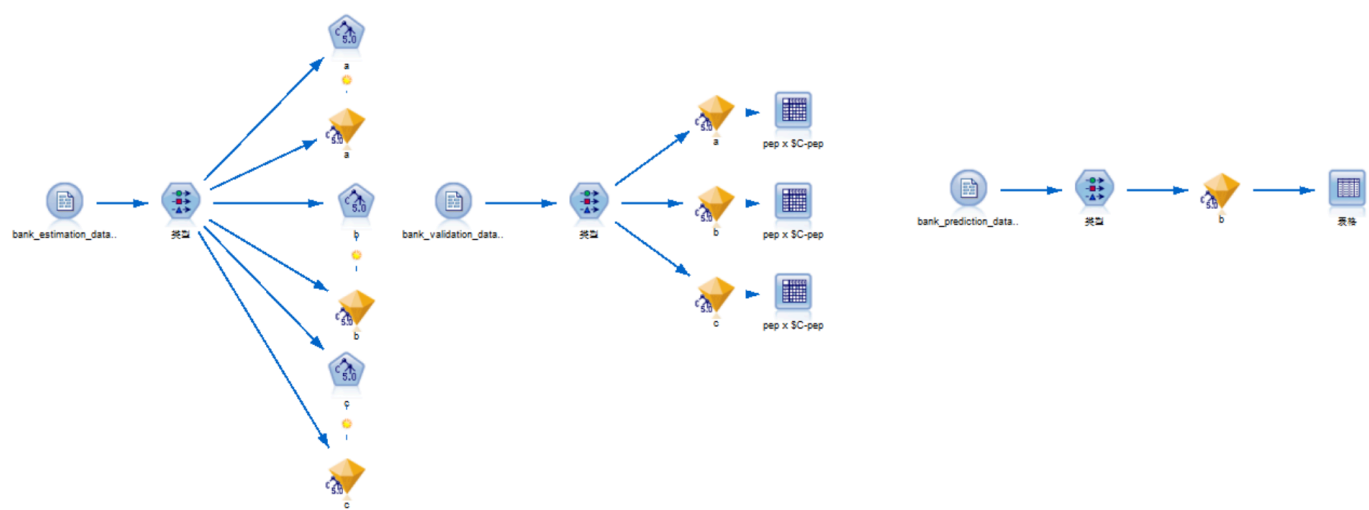
### 3

1.  $P(\text{抽烟}) = P(\text{抽烟}|\text{研究生}) \times P(\text{研究生}) + P(\text{抽烟}|\text{本科生}) \times P(\text{本科生})$   
 $P(\text{抽烟}) = 0.23 \times 0.2 + 0.15 \times 0.8 = 0.046 + 0.12 = 0.166$   
 $P(\text{研究生}|\text{抽烟}) = [P(\text{抽烟}|\text{研究生}) \times P(\text{研究生})] / P(\text{抽烟}) = 0.23 \times 0.2 / 0.166 = 0.046 / 0.166 \approx 0.276$
2.  $P(\text{本科生}) = 0.8$ ,  $P(\text{研究生}) = 0.2$ ,  $P(\text{本科生}) > P(\text{研究生})$   
 故随机选择的学生更可能是本科生
3. 由独立性可得:  
 $P(\text{抽烟且住宿舍}|\text{研究生}) = 0.23 \times 0.3 = 0.069$   
 $P(\text{抽烟且住宿舍}|\text{本科生}) = 0.15 \times 0.1 = 0.015$   
 则有:  
 $P(\text{抽烟且住宿舍}) = 0.069 \times 0.2 + 0.015 \times 0.8 = 0.0138 + 0.012 = 0.0258$

$P(\text{研究生}|\text{抽烟且住宿舍}) = (0.069 \times 0.2) / 0.0258 = 0.0138 / 0.0258 \approx 0.535$   
 $P(\text{本科生}|\text{抽烟且住宿舍}) = 1 - 0.535 = 0.465$   
 $P(\text{研究生}|\text{抽烟且住宿舍}) > P(\text{本科生}|\text{抽烟且住宿舍})$   
故一个抽烟且住宿舍的学生更可能是研究生

上机作业1

1. 建模过程如图:



混淆矩阵见压缩包

2. 根据三个模型的混淆矩阵分析，应选择决策树模型b，理由如下：
- i. 根据准确率判断，b(84.5%) 和 c(84%) 的表现远优于 a(71.5%)
  - ii. 相比 c，b 的精确率与F1分数更优，说明其误判更少，且其在精确率与召回率的平衡性表现更佳
- 3.

age	sex	region	income	married	children	car	save_act	current_act	mortgage	RECOMMEND PEP (Y/N)
22	0	1	14000	0	3	0	1	1	0	0
34	1	0	33000	0	0	1	1	0	0	1
47	0	0	16700	1	1	0	1	1	0	0
54	1	1	43400	1	1	1	1	1	0	1
65	1	2	60000	1	1	0	1	1	0	1
37	0	0	27700	0	1	1	0	0	0	1
44	0	0	38784	1	0	0	1	1	0	0
20	1	0	10200	1	0	0	1	1	1	0
46	0	0	22000	1	1	1	1	0	1	1
40	1	1	37400	1	2	0	1	1	0	1

上机作业2

混淆矩阵见压缩包