# Data Mining: Homework 2

Illusionna 2025.............4 Artificial Intelligence School

20:48, Wednesday 26th November, 2025

## Contents

# I Written Assignment

## 1 ID3 Algorithm

Table 1: The customer data overview (Q1)

| Customer ID | Gender | Car Type | Shirt Size | Class |
|---|---|---|---|---|
| 1 | M | Family | Small | C0 |
| 2 | M | Sports | Medium | C0 |
| 3 | M | Sports | Medium | C0 |
| 4 | M | Sports | Large | C0 |
| 5 | M | Sports | Extra Large | C0 |
| 6 | M | Sports | Extra Large | C0 |
| 7 | F | Sports | Small | C0 |
| 8 | F | Sports | Small | C0 |
| 9 | F | Sports | Medium | C0 |
| 10 | F | Luxury | Large | C0 |
| 11 | M | Family | Large | C1 |
| 12 | M | Family | Extra Large | C1 |
| 13 | M | Family | Medium | C1 |
| 14 | M | Luxury | Extra Large | C1 |
| 15 | F | Luxury | Small | C1 |
| 16 | F | Luxury | Small | C1 |
| 17 | F | Luxury | Medium | C1 |
| 18 | F | Luxury | Medium | C1 |
| 19 | F | Luxury | Medium | C1 |
| 20 | F | Luxury | Large | C1 |

### (a) Compute the information gain

*Compute the Information Gain for Gender, Car Type and Shirt Size.*

$$IG(S) = \text{Entropy}(D) - \sum_{n \in S} \frac{|D_n|}{|D|} \cdot \text{Entropy}(D_n)$$

$$\text{Entropy}(X) = -\sum_{n=1}^{N} p_n \log_2 p_n$$

Among: $D$ denotes the class of dataset, $D_n$ denotes a certain attribute of sub-dataset. $|\sim|$ denotes the element count.

$$\text{C0} = \text{C1} = 10 \implies p(\text{C0}) = p(\text{C1}) = 0.5 \implies \text{Entropy}(\text{Class}) = -0.5 \log_2 0.5 \times 2 = 1$$

Table 2: Gender Attribute

| Gender | C0 | C1 | Entropy |
|---|---|---|---|
| M | 6 | 4 | 0.9709505944546686 |
| F | 4 | 6 | 0.9709505944546686 |

$$IG(\text{Gender}) = 1 - \frac{6+4}{20} \times 0.9709505944546686 \times 2 = 0.029049405545533142$$

Table 3: Car Type Attribute

| Gender | C0 | C1 | Entropy |
|---|---|---|---|
| Family | 1 | 3 | 0.8112781244591328 |
| Sports | 8 | 0 | 0 |
| Luxury | 1 | 7 | 0.5435644431995964 |

$$IG(\text{Car Type}) = 1 - \frac{4}{20} \times 0.8112781244591328 - \frac{8}{20} \times 0.5435644431995964 = 0.620318597828335$$

Table 4: Shirt Size Attribute

| Shirt Size | C0 | C1 | Entropy |
|---|---|---|---|
| Small | 3 | 2 | $-\frac{3}{3+2}\log_2\left(\frac{3}{3+2}\right) - \frac{2}{3+2}\log_2\left(\frac{2}{3+2}\right) \approx 0.9709505944546686$ |
| Medium | 3 | 4 | $-\frac{3}{3+4}\log_2\left(\frac{3}{3+4}\right) - \frac{4}{3+4}\log_2\left(\frac{4}{3+4}\right) \approx 0.9852281360342515$ |
| Large | 2 | 2 | $-\frac{2}{2+2}\log_2\left(\frac{2}{2+2}\right) - \frac{2}{2+2}\log_2\left(\frac{2}{2+2}\right) = 1$ |
| Extra Large | 2 | 2 | $-\frac{2}{2+2}\log_2\left(\frac{2}{2+2}\right) - \frac{2}{2+2}\log_2\left(\frac{2}{2+2}\right) = 1$ |

$$\begin{aligned}
IG(\text{Shirt Size}) &= \text{Entropy}(\text{Class}) - \sum_{n \in \text{Shirt Size}} \frac{|D_n|}{|\text{Class}|} \cdot \text{Entropy}(D_n) \\
&= E(\text{Class}) - \frac{|\text{Small}|}{|\text{Class}|} E(\text{Small}) - \frac{|\text{Medium}|}{|\text{Class}|} E(\text{Medium}) \\
&\quad - \frac{|\text{Large}|}{|\text{Class}|} E(\text{Large}) - \frac{|\text{Extra Large}|}{|\text{Class}|} E(\text{Extra Large}) \\
&= 1 - \frac{3+2}{20} \times 0.9709505944546686 - \frac{3+4}{20} \times 0.9852281360342515 \\
&\quad - \frac{2+2}{20} \times 1 - \frac{2+2}{20} \times 1 \\
&= 0.012432503774344739
\end{aligned}$$

### (b) Construct a decision tree

*Construct a decision tree with Information Gain.*

**Step 1**

Use "Car Type" as the selection attribute to partition the current dataset.

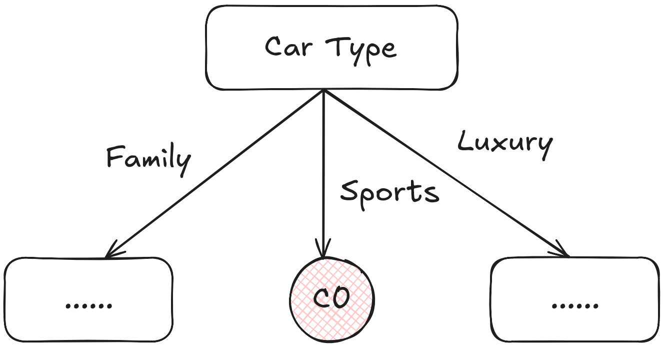


Figure 1: Decision tree step 1

**Step 2**

**For the dataset partitioned by "Car Type = Family", we calculate the information gain in this subset as follows:**

Table 5: Car Type = Family

| Customer ID | Gender | Car Type | Shirt Size | Class |
|---|---|---|---|---|
| 1 | M | Family | Small | C0 |
| 11 | M | Family | Large | C1 |
| 12 | M | Family | Extra Large | C1 |
| 13 | M | Family | Medium | C1 |

$$\text{Entropy(Class)} = -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{3}{4}\log_2\left(\frac{3}{4}\right) \approx 0.8112781244591328$$

Table 6: Gender Attribute (Car Type = Family)

| Gender | C0 | C1 | Entropy |
|---|---|---|---|
| M | 1 | 3 | 0.8112781244591328 |
| F | 0 | 0 | 0 |

$$IG(\text{Gender}) = 0.8112781244591328 - \frac{1+3}{4} \times (0.8112781244591328) = {\color{red}0}$$

Table 7: Shirt Size Attribute (Car Type = Family)

| Shirt Size | C0 | C1 | Entropy |
|---|---|---|---|
| Small | 1 | 0 | 0 |
| Medium | 0 | 1 | 0 |
| Large | 0 | 1 | 0 |
| Extra Large | 0 | 1 | 0 |

$$IG(\text{Shirt Size}) = 0.8112781244591328 - \frac{1+0}{4} \times 0 - \frac{0+1}{4} \times 0 \times 3 = 0.8112781244591328$$

$$IG(\text{Shirt Size}) > IG(\text{Gender})$$

Select "Shirt Size" as partition attribute. Note that all information entropy in "Shirt Size" is 0 and no branching is required.

**For the dataset partitioned by "Car Type = Luxury", we calculate the information gain in this subset as follows:**

Table 8: Car Type = Luxury

| Customer ID | Gender | Car Type | Shirt Size | Class |
|---|---|---|---|---|
| 10 | F | Luxury | Large | C0 |
| 14 | M | Luxury | Extra Large | C1 |
| 15 | F | Luxury | Small | C1 |
| 16 | F | Luxury | Small | C1 |
| 17 | F | Luxury | Medium | C1 |
| 18 | F | Luxury | Medium | C1 |
| 19 | F | Luxury | Medium | C1 |
| 20 | F | Luxury | Large | C1 |

$$\text{Entropy}(\text{Class}) = -\frac{1}{8}\log_2\left(\frac{1}{8}\right) - \frac{7}{8}\log_2\left(\frac{7}{8}\right) \approx 0.5435644431995964$$

Table 9: Gender Attribute (Car Type = Luxury)

| Gender | C0 | C1 | Entropy |
|---|---|---|---|
| M | 0 | 1 | 0 |
| F | 1 | 6 | 0.5916727785823275 |

$$IG(\text{Gender}) = 0.5435644431995964 - \frac{0+1}{8} \times 0 - \frac{1+6}{8} \times 0.5916727785823275$$
$$\approx 0.025850761940059863$$

Table 10: Shirt Size Attribute (Car Type = Luxury)

| Shirt Size | C0 | C1 | Entropy |
|---|---|---|---|
| Small | 0 | 2 | 0 |
| Medium | 0 | 3 | 0 |
| Large | 1 | 1 | $1^{\dagger}$ |
| Extra Large | 0 | 1 | 0 |

$$IG(\text{Shirt Size}) = 0.5435644431995964 - \frac{0+2}{8} \times 0 - \frac{0+3}{8} \times 0 - \frac{1+1}{8} \times 1 - \frac{0+1}{8} \times 0$$
$$= 0.2935644431995964$$

$$IG(\text{Shirt Size}) > IG(\text{Gender})$$

Select "Shirt Size" as partition attribute. Note that "Large" information entropy in "Shirt Size" is 1† and branching is still required.

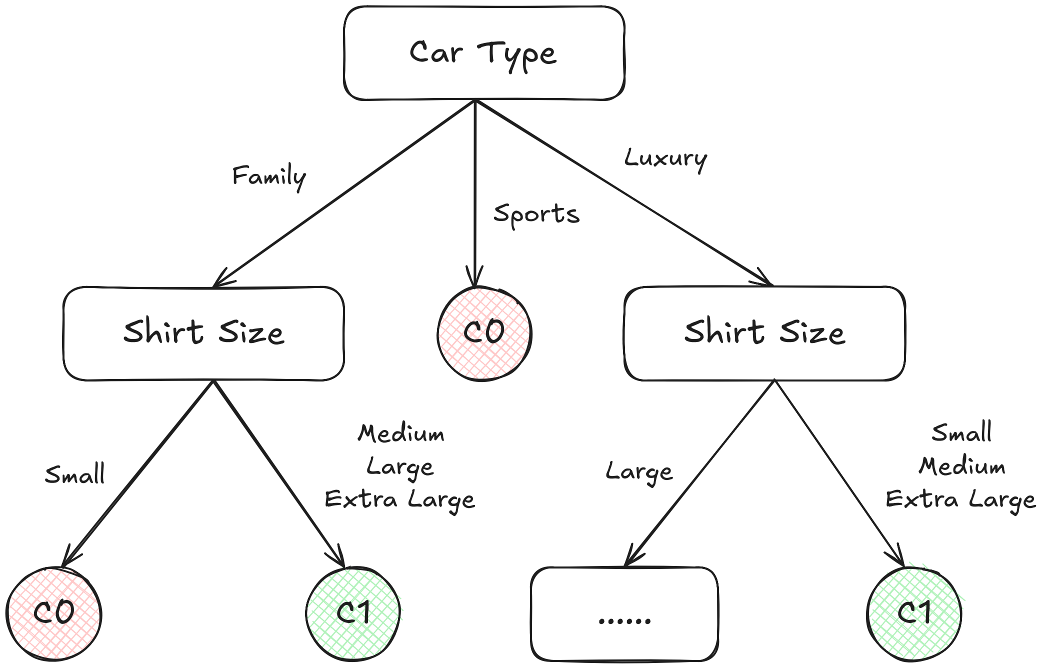


Figure 2: Decision tree step 2

**Step 3**

**For the dataset partitioned by "Car Type = Family" & "Shirt Size = Large", the "Gender" attribute is only available here:**

Table 11: Car Type = Luxury & Shirt Size = Large

| Customer ID | Gender | Car Type | Shirt Size | Class |
|---|---|---|---|---|
| 10 | F | Luxury | Large | C0 |
| 20 | F | Luxury | Large | C1 |

Since in the "Gender" attribute, one belongs to C0 and the other to C1, it is a tie. The phenomenon of branch conflicts caused by data inconsistency has emerged. The majority voting strategy is applied to determine the node. It is important to trace back to the parent node. The prior probability of parent node (Car Type = Luxury) is:

$$C0 = 1 \quad C1 = 7$$

Because C1 in parent node has an absolute advantage, the key of tie-breaking is choosing C1 when the child node (Car Type = Luxury & Shirt Size = Large) is tie.

After pruning, the final decision tree is given:

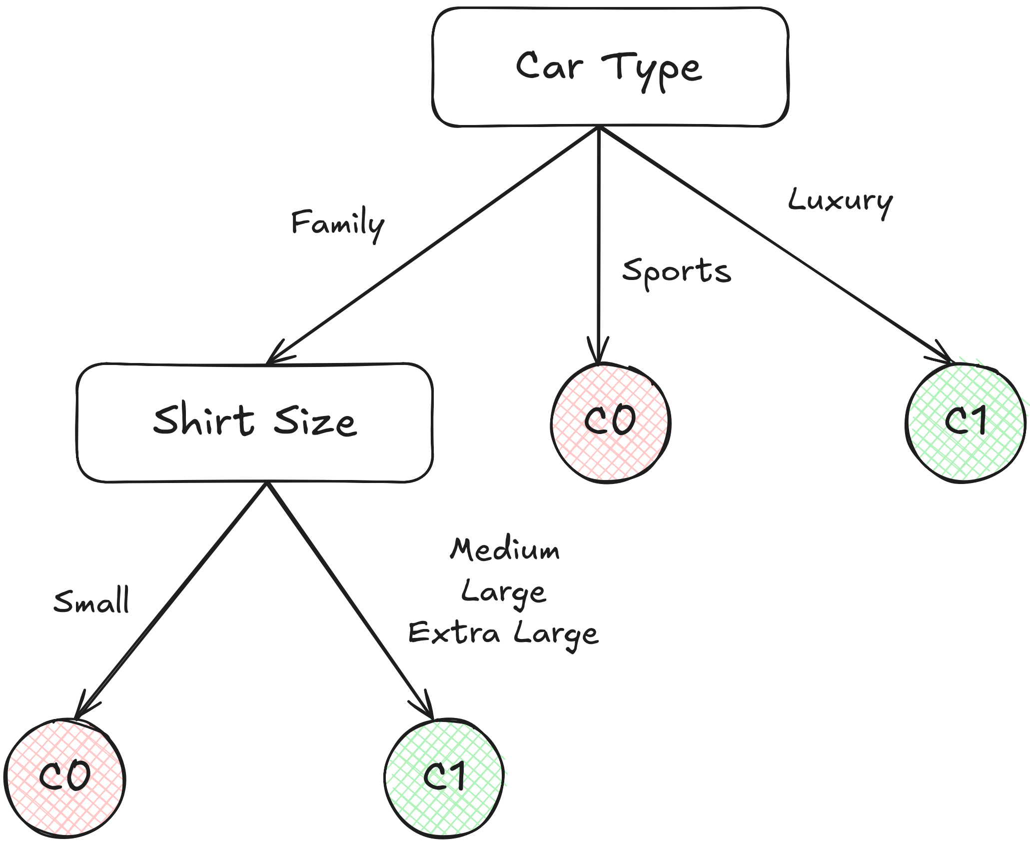


Figure 3: Decision tree step 3

## 2 Neural Network

### (a) Design a multilayer feed-forward neural network

*Design a multilayer feed-forward neural network (one hidden layer) for the data set in Q1 (customer data overview). Label the nodes in the input and output layers.*

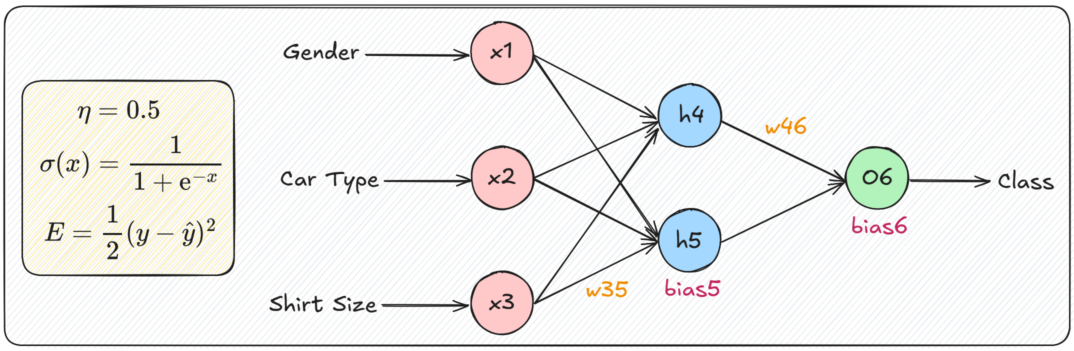


Figure 4: Multilayer feed-forward neural network

$$\text{Gender}\begin{cases}\text{M} = 0\\ \text{F} = 1\end{cases} \qquad \text{Car Type}\begin{cases}\text{Family} = 0\\ \text{Sports} = 1\\ \text{Luxury} = 2\end{cases} \qquad \text{Shirt Size}\begin{cases}\text{Small} = 0\\ \text{Medium} = 1\\ \text{Large} = 2\\ \text{Extra Large} = 3\end{cases}$$

The output of neural network is Class, denoting C0 or C1.

### (b) Update the weight values after one iteration of the back propagation

*Using the neural network obtained above, show the weight values after one iteration of the back propagation algorithm, given the training instance "(M, Family, Small, C0)". Indicate your initial weight values and biases and the learning rate used.*

Initialize hyper-parameter $\eta = 0.5$ and random parameters as follows:

Table 12: Random initialized parameters

| Parameter | Value |
|---|---|
| $\omega_{14}$ | 0.1 |
| $\omega_{24}$ | 0.2 |
| $\omega_{34}$ | $-0.1$ |
| $\omega_{15}$ | $-0.1$ |
| $\omega_{25}$ | 0.1 |
| $\omega_{35}$ | 0.05 |
| $\omega_{46}$ | 0.4 |
| $\omega_{56}$ | $-0.3$ |
| $b_4$ | 0.35 |
| $b_5$ | $-0.35$ |
| $b_6$ | 0.1 |

**Feed forward**

$$
\begin{aligned}
z_4 &= [\omega_{14},\ \omega_{24},\ \omega_{34}] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b_4 = [0.1,\ 0.2,\ -0.1] \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0.35 = 0.35 \\
a_4 &= \sigma(z_4) = \frac{1}{1+\mathrm{e}^{-z_4}} \approx 0.5866175789173301 \\
z_5 &= [\omega_{15},\ \omega_{25},\ \omega_{35}] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + b_5 = [-0.1,\ 0.1,\ 0.05] \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - 0.35 = -0.35 \\
a_5 &= \sigma(z_5) = \frac{1}{1+\mathrm{e}^{-z_5}} =\approx 0.41338242108267 \\
z_6 &= [\omega_{46},\ \omega_{56}] \begin{bmatrix} a_4 \\ a_5 \end{bmatrix} + b_6 = [0.4,\ -0.3] \begin{bmatrix} 0.5866175789173301 \\ 0.41338242108267 \end{bmatrix} + 0.1 = 0.21063230524213108 \\
a_6 &= \sigma(z_6) = \frac{1}{1+\mathrm{e}^{-z_6}} \approx 0.5524642506473584
\end{aligned}
$$

$$
E = \frac{1}{2}(y-\hat{y})^2 = \frac{1}{2}(0-a_6)^2 = \frac{1}{2}(0-0.5524642506473584)^2 \approx 0.1526083741216736
$$

**Back-propagation**

$$\mathcal{L}_6 = \frac{\partial E}{\partial z_6} = \frac{\partial E}{\partial a_6} \cdot \frac{\partial a_6}{\partial z_6} = \left.\frac{\mathrm{d}\left[\frac{1}{2}(y-\hat{y})^2\right]}{\mathrm{d}\hat{y}}\right|_{\hat{y}=a_6} \cdot \left.\frac{\mathrm{d}\left[\frac{1}{1+\mathrm{e}^{-x}}\right]}{\mathrm{d}x}\right|_{x=z_6} = (a_6 - y)a_6(1-a_6)$$

$$= (0.5524642506473584 - 0) \times 0.5524642506473584 \times (1 - 0.5524642506473584)$$
$$\approx 0.13659540614006296$$

According to:

$$z_6 = [\omega_{46},\ \omega_{56}] \begin{bmatrix} a_4 \\ a_5 \end{bmatrix} + b_6 \simeq \begin{cases} f(\boldsymbol{x}) = \sum \boldsymbol{\omega x} + b \implies f'(x) = \omega \\ f(\boldsymbol{\omega}) = \sum \boldsymbol{x\omega} + b \implies f'(\omega) = x \\ f(b) = b + \sum \boldsymbol{\omega x} \implies f'(b) \equiv 1 \end{cases}$$

Then:

$$\mathcal{L}_5 = \frac{\partial E}{\partial z_5} = \frac{\partial E}{\partial z_6} \cdot \frac{\partial z_6}{\partial a_5} \cdot \frac{\partial a_5}{\partial z_5} = \mathcal{L}_6 \times \omega_{56} \times a_5(1-a_5)$$
$$= 0.13659540614006296 \times (-0.3) \times 0.41338242108267 \times (1 - 0.41338242108267)$$
$$\approx -0.009937209048301705$$

$$\mathcal{L}_4 = \frac{\partial E}{\partial z_4} = \frac{\partial E}{\partial z_6} \cdot \frac{\partial z_6}{\partial a_4} \cdot \frac{\partial a_4}{\partial z_4} = \mathcal{L}_6 \times \omega_{46} \times a_4(1-a_4)$$
$$= 0.13659540614006296 \times 0.4 \times 0.5866175789173301 \times (1 - 0.5866175789173301)$$
$$\approx 0.013249612064402273$$

Therefore:

$$\omega'_{46} = \omega_{46} - \eta\frac{\partial E}{\partial \omega_{46}} = \omega_{46} - \eta \cdot \frac{\partial E}{\partial z_6} \cdot \frac{\partial z_6}{\partial \omega_{46}} = \omega_{46} - \eta\mathcal{L}_6 a_4$$
$$= 0.4 - 0.5 \times 0.13659540614006296 \times 0.5866175789173301$$
$$= 0.35993536677944343$$

$$\omega'_{56} = \omega_{56} - \eta\frac{\partial E}{\partial \omega_{56}} = \omega_{56} - \eta \cdot \frac{\partial E}{\partial z_6} \cdot \frac{\partial z_6}{\partial \omega_{56}} = \omega_{56} - \eta\mathcal{L}_6 a_5$$
$$= -0.3 - 0.5 \times 0.13659540614006296 \times 0.41338242108267$$
$$= -0.3282330698494749$$

$$\omega'_{14} = \omega_{14} - \eta\frac{\partial E}{\partial \omega_{14}} = \omega_{14} - \eta \cdot \frac{\partial E}{\partial z_4} \cdot \frac{\partial z_4}{\partial \omega_{14}} = \omega_{14} - \eta\mathcal{L}_4 x_1 = \omega_{14}$$
$$\omega'_{24} = \omega_{24} - \eta\frac{\partial E}{\partial \omega_{24}} = \omega_{24} - \eta \cdot \frac{\partial E}{\partial z_4} \cdot \frac{\partial z_4}{\partial \omega_{24}} = \omega_{24} - \eta\mathcal{L}_4 x_2 = \omega_{24}$$
$$\omega'_{34} = \omega_{34} - \eta\frac{\partial E}{\partial \omega_{34}} = \omega_{34} - \eta \cdot \frac{\partial E}{\partial z_4} \cdot \frac{\partial z_4}{\partial \omega_{34}} = \omega_{34} - \eta\mathcal{L}_4 x_3 = \omega_{34}$$

$$\omega'_{15} = \omega_{15} - \eta \frac{\partial E}{\partial \omega_{15}} = \omega_{15} - \eta \cdot \frac{\partial E}{\partial z_5} \cdot \frac{\partial z_5}{\partial \omega_{15}} = \omega_{15} - \eta \mathcal{L}_5 x_1 = \omega_{15}$$
$$\omega'_{25} = \omega_{25} - \eta \frac{\partial E}{\partial \omega_{25}} = \omega_{25} - \eta \cdot \frac{\partial E}{\partial z_5} \cdot \frac{\partial z_5}{\partial \omega_{25}} = \omega_{25} - \eta \mathcal{L}_5 x_2 = \omega_{25}$$
$$\omega'_{35} = \omega_{35} - \eta \frac{\partial E}{\partial \omega_{35}} = \omega_{35} - \eta \cdot \frac{\partial E}{\partial z_5} \cdot \frac{\partial z_5}{\partial \omega_{35}} = \omega_{35} - \eta \mathcal{L}_5 x_3 = \omega_{35}$$

And these biases:

$$\begin{aligned} b'_6 &= b_6 - \eta \frac{\partial E}{\partial b_6} = b_6 - \eta \cdot \frac{\partial E}{\partial z_6} \cdot \frac{\partial z_6}{\partial b_6} = b_6 - \eta \mathcal{L}_6 \\ &= 0.1 - 0.5 \times 0.13659540614006296 \\ &\approx 0.031702296929968524 \end{aligned}$$

$$\begin{aligned} b'_5 &= b_5 - \eta \frac{\partial E}{\partial b_5} = b_5 - \eta \cdot \frac{\partial E}{\partial z_5} \cdot \frac{\partial z_5}{\partial b_5} = b_5 - \eta \mathcal{L}_5 \\ &= -0.35 - 0.5 \times (-0.009937209048301705) \\ &\approx -0.3450313954758491 \end{aligned}$$

$$\begin{aligned} b'_4 &= b_4 - \eta \frac{\partial E}{\partial b_4} = b_4 - \eta \cdot \frac{\partial E}{\partial z_4} \cdot \frac{\partial z_4}{\partial b_4} = b_4 - \eta \mathcal{L}_4 \\ &= 0.35 - 0.5 \times 0.013249612064402273 \\ &\approx 0.34337519396779886 \end{aligned}$$

Summarize:

Table 13: Parameters of one iteration of BP algorithm

| Parameter | Update |
|---|---|
| $\omega_{14}$ | $0.1 \to 0.1$ |
| $\omega_{24}$ | $0.2 \to 0.2$ |
| $\omega_{34}$ | $-0.1 \to -0.1$ |
| $\omega_{15}$ | $-0.1 \to -0.1$ |
| $\omega_{25}$ | $0.1 \to 0.1$ |
| $\omega_{35}$ | $0.05 \to 0.05$ |
| $\omega_{46}$ | $0.4 \to 0.35993536677944343$ |
| $\omega_{56}$ | $-0.3 \to -0.3282330698494749$ |
| $b_4$ | $0.35 \to 0.34337519396779886$ |
| $b_5$ | $-0.35 \to -0.3450313954758491$ |
| $b_6$ | $0.1 \to 0.031702296929968524$ |

## 3 Naïve Bayesian Classifier

*Classify the unknown sample Z based on the training data set in Q1 (customer data overview): Z = (Gender = M, Car Type = Sports, Shirt Size = Small). What would a naïve Bayesian classifier classify Z?*

$$p(\text{Class} = \text{C0}) = p(\text{Class} = \text{C1}) = \frac{10}{20} = \frac{1}{2}$$
$$p(\text{Gender} = \text{M}|\text{Class} = \text{C0}) = \frac{6}{6+4} = \frac{3}{5}$$
$$p(\text{Gender} = \text{M}|\text{Class} = \text{C1}) = \frac{4}{6+4} = \frac{2}{5}$$
$$p(\text{Car Type} = \text{Sports}|\text{Class} = \text{C0}) = \frac{8}{1+8+1} = \frac{4}{5}$$
$$p(\text{Car Type} = \text{Sports}|\text{Class} = \text{C1}) = \frac{0}{3+0+7} = 0$$
$$p(\text{Shirt Size} = \text{Small}|\text{Class} = \text{C0}) = \frac{3}{3+3+2+2} = \frac{3}{10}$$
$$p(\text{Shirt Size} = \text{Small}|\text{Class} = \text{C1}) = \frac{2}{3+3+2+2} = \frac{1}{5}$$

$$\begin{aligned} p(\text{Z}|\text{Class} = \text{C0}) = {} & p(\text{Gender} = \text{M}|\text{Class} = \text{C0}) \\ & \times p(\text{Car Type} = \text{Sports}|\text{Class} = \text{C0}) \times p(\text{Shirt Size} = \text{Small}|\text{Class} = \text{C0}) \\ = {} & \frac{3}{5} \times \frac{4}{5} \times \frac{3}{10} \end{aligned}$$

$$\begin{aligned} p(\text{Z}|\text{Class} = \text{C1}) = {} & p(\text{Gender} = \text{M}|\text{Class} = \text{C1}) \\ & \times p(\text{Car Type} = \text{Sports}|\text{Class} = \text{C1}) \times p(\text{Shirt Size} = \text{Small}|\text{Class} = \text{C1}) \\ = {} & \frac{2}{5} \times 0 \times \frac{1}{5} \end{aligned}$$

$$p(\text{Z}|\text{Class} = \text{C0})p(\text{Class} = \text{C0}) > p(\text{Z}|\text{Class} = \text{C1})p(\text{Class} = \text{C1}) \implies \text{Z} \in \text{C0}$$

## 4 Apriori Algorithm

*Consider the data set shown in table (min_sup = 60%, min_conf=70%).*

Table 14: Example of market basket transactions

| TID | Items-bought |
|---|---|
| T1 | {A, D, B, C} |
| T2 | {D, A, C, E, B} |
| T3 | {A, B, E} |
| T4 | {A, B, D} |

### (a) Find all frequent itemsets

*Find all frequent itemsets using Apriori by treating each transaction ID as a market basket.*

$$\text{min\_count} = \lceil \text{min\_sup} \times n \rceil = \lceil 60\% \times 4 \rceil = 3$$

1st scan:

Table 15: Frequent itemset 1

| Itemset | TID | Count | Pass |
|---|---|---|---|
| {A} | T1, T2, T3, T4 | 4 | ✓ |
| {B} | T1, T2, T3, T4 | 4 | ✓ |
| {C} | T1, T2 | 2 | ✗ |
| {D} | T1, T2, T4 | 3 | ✓ |
| {E} | T2, T3 | 2 | ✗ |

2nd scan:

Table 16: Frequent itemset 2

| Itemset | TID | Count | Pass |
|---|---|---|---|
| {A, B} | T1, T2, T3, T4 | 4 | ✓ |
| {A, D} | T1, T2, T4 | 3 | ✓ |
| {B, D} | T1, T2, T4 | 3 | ✓ |

1rd scan:

Table 17: Frequent itemset 3

| Itemset | TID | Count | Pass |
|---|---|---|---|
| {A, B, D} | T1, T2, T4 | 3 | ✓ |

All the frequent itemsets are: {A}, {B}, {D}, {A, B}, {A, D}, {B, D}, {A, B, D}.

### (b) Compute the confidence for the association rules

*Use the results in part (a) to compute the confidence for the association rules $\{a,\ b\} \rightarrow \{c\}$ and $\{c\} \rightarrow \{a,\ b\}$. Is confidence a symmetric measure?*

$$\text{Suppor}(X \rightarrow Y) = \text{Suppor}(X \bigcup Y)$$

$$\text{Confidence}\left(\{A,\ B\} \rightarrow \{C\}\right) = \frac{\text{Support}\left(\{A,\ B\} \bigcup \{C\}\right)}{\text{Support}\left(\{A,\ B\}\right)} = \frac{\frac{2}{n}}{\frac{4}{n}} = 50\%$$

$$\text{Confidence}\left(\{C\} \rightarrow \{A,\ B\}\right) = \frac{\text{Support}\left(\{C\} \bigcup \{A,\ B\}\right)}{\text{Support}\left(\{C\}\right)} = \frac{\frac{2}{n}}{\frac{2}{n}} = 100\%$$

$$\text{Confidence}\left(\{A,\ B\} \rightarrow \{C\}\right) \neq \text{Confidence}\left(\{C\} \rightarrow \{A,\ B\}\right)$$

Since here we have, the confidence is an asymmetric measure which depends on precondition.

### (c) List all of the strong association rules

*List all of the strong association rules (with support s and confidence c) matching the following metarule, where X is a variable representing customers, and* $\text{item}_i$ *denotes variables representing items (e.g. "A", "B", etc.):*

$$\forall x \in \text{transactions}, \quad \text{buys}(X,\ \text{item}_1) \wedge \text{buys}(X,\ \text{item}_2) => \text{buys}(X,\ \text{item}_3), \quad [s,\ c]$$

Construct 3 association rules with "{A, B, D}":

$$\begin{aligned} \text{rule 1}: &\quad \{A,\ B\} \to \{D\} \\ \text{rule 2}: &\quad \{A,\ D\} \to \{B\} \\ \text{rule 3}: &\quad \{B,\ D\} \to \{A\} \end{aligned}$$

$$\text{Suppor}(\{A,\ B\} \to \{D\}) = \text{Suppor}(\{A,\ D\} \to \{B\}) = \text{Suppor}(\{B,\ D\} \to \{A\}) = 75\%$$

$$\text{Confidence}\left(\{A,\ B\} \to \{D\}\right) = \frac{\text{Support}\left(\{A,\ B\} \bigcup \{D\}\right)}{\text{Support}\left(\{A,\ B\}\right)} = \frac{3}{4} = 75\%$$
$$\text{Confidence}\left(\{A,\ D\} \to \{B\}\right) = \frac{\text{Support}\left(\{A,\ D\} \bigcup \{B\}\right)}{\text{Support}\left(\{A,\ D\}\right)} = \frac{3}{3} = 100\%$$
$$\text{Confidence}\left(\{B,\ D\} \to \{A\}\right) = \frac{\text{Support}\left(\{B,\ D\} \bigcup \{A\}\right)}{\text{Support}\left(\{B,\ D\}\right)} = \frac{3}{3} = 100\%$$

Since all the confidence values of association rule are higher than those of min_conf, the 3 rules ($\{A,\ B\} \to \{D\}$, $\{A,\ D\} \to \{B\}$, $\{B,\ D\} \to \{A\}$) are strong association rules.

# II Lab Assignment

## 1 C5.0 Algorithm for Transactions

*Assume this supermarket would like to promote milk. Use the data in "transactions" as training data to build a decision tree (C5.0 algorithm) model to predict whether the customer would buy pasta or not.*

### (a) A figure showing decision tree

*Build a decision tree using data set "transactions" that predicts milk as a function of the other fields. Set the "type" of each field to "Flag", set the "direction" of "pasta" as "out", set the "type" of COD as "Typeless", select "Expert" and set the "pruning severity" to 65, and set the "minimum records per child branch" to be 95.*

***Hand-in: a figure showing your tree.***

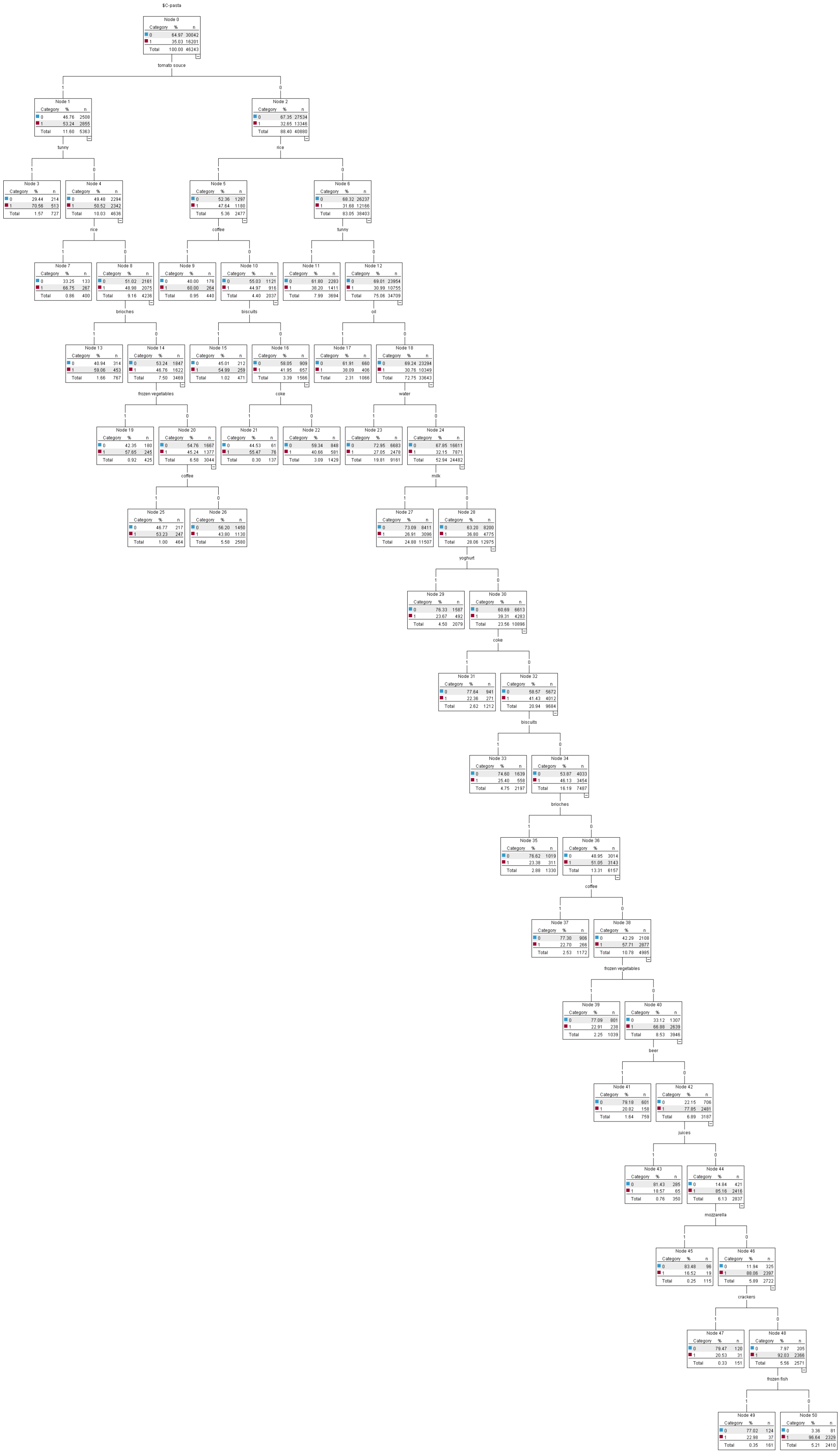

Figure 5: C5.0 decision tree

## (b) Prediction for each of the 20 customers

*Use the model (the full tree generated by Clementine in step 1 above) to make a prediction for each of the 20 customers in the "rollout" data to determine whether the customer would buy pasta.*

***Hand-in: your prediction for each of the 20 customers.***

Table (21 fields, 20 records) #2

File Edit Generate

| | milk | water | biscuits | coffee | brioches | yoghurt | frozen vegetables | tunny | beer | tomato souce | coke | rice | juices | crackers | oil | frozen fish | ice cream | mozzarella | tinned meat | $C-pasta | $CC-pasta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0.731 |
| 2 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.746 |
| 3 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0.618 |
| 4 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0.729 |
| 5 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0.618 |
| 6 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.776 |
| 7 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.763 |
| 8 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.729 |
| 9 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0.618 |
| 10 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0.600 |
| 11 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.746 |
| 12 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.731 |
| 13 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.729 |
| 14 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.729 |
| 15 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.593 |
| 16 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.729 |
| 17 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.554 |
| 18 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0.593 |
| 19 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.731 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.562 |

Figure 6: Prediction for each of the 20 customersr

## (c) Rules for positive prediction of pasta purchase

***Hand-in: rules for positive (yes) prediction of pasta purchase identified from the decision tree (up to the fifth level. The root is considered as level 1).***

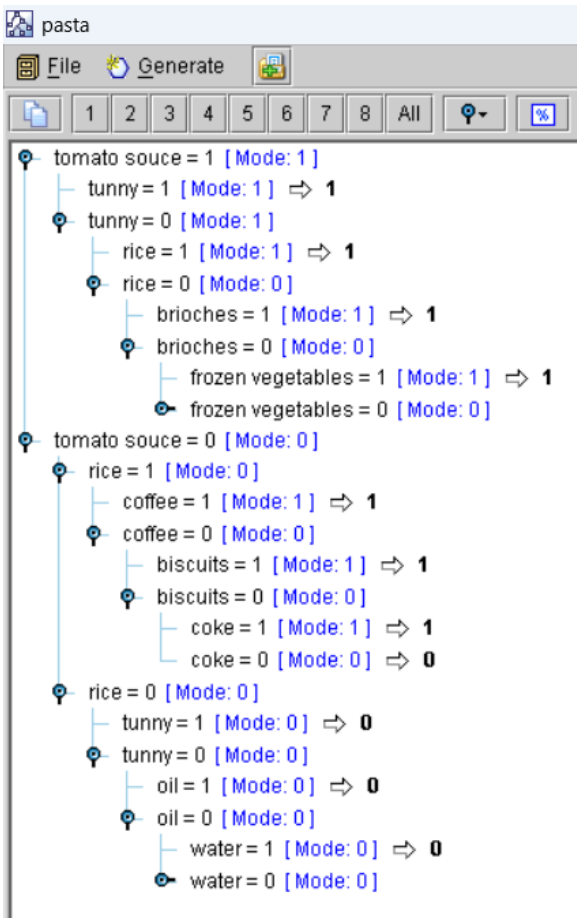


Figure 7: Rules for positive prediction of pasta purchase

```
Rules for 1 - contains 7 rule(s)
    Rule 1 for 1
        if tomato souce = 1
        and tunny = 1
        then 1
    Rule 2 for 1
        if tomato souce = 1
        and tunny = 0
        and rice = 1
        then 1
    Rule 3 for 1
        if tomato souce = 1
        and tunny = 0
        and rice = 0
        and brioches = 1
        then 1
    Rule 4 for 1
        if tomato souce = 1
        and tunny = 0
        and rice = 0
        and brioches = 0
        and frozen vegetables = 1
        then 1
    Rule 6 for 1
        if tomato souce = 0
        and rice = 1
        and coffee = 1
        then 1
    Rule 7 for 1
        if tomato souce = 0
        and rice = 1
        and coffee = 0
        and biscuits = 1
        then 1
    Rule 8 for 1
        if tomato souce = 0
        and rice = 1
        and coffee = 0
        and biscuits = 0
        and coke = 1
        then 1
```

## 2 C5.0 Algorithm for Bank PEP

*Each record is a customer description where the "pep" field indicates whether or not that customer bought a PEP. For other existing customers in the database, we would like to see if PEP should be RECOMMENDED to the customers in the roll-out data.*

*The firm decides to use decision tree to build the models for PEP recommendation. Develop a decision tree model using the estimation data. For building this model, you are expected to use the following steps.*

*Using the "bank-estimation-data", estimate the decision tree that predicts pep as a function of the other variables. Select "Expert" and set "pruning severity" at 70. Set the "Type" of pep as "Flag" and the "Direction" as "out". Build decision trees using three options "Minimum records per child branch" values being (a) 56, (b) 15 and (c) 10, not selecting "use global pruning".*

### (a) Confusion matrix for different strategies

***Hand in: the confusion matrix for (a), (b) and (c) on the validation data.***

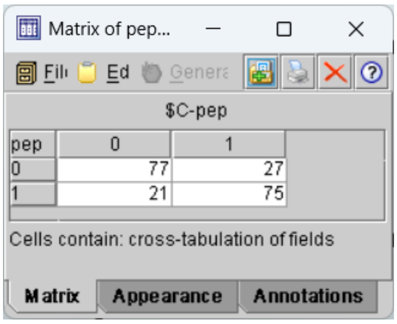


Figure 8: Minimum records per child branch (a) 56

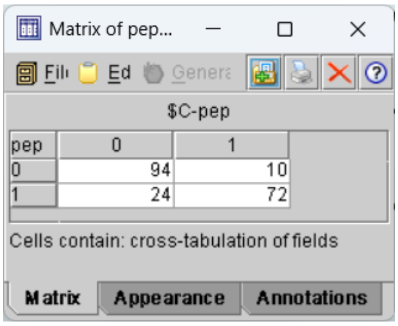


Figure 9: Minimum records per child branch (b) 15

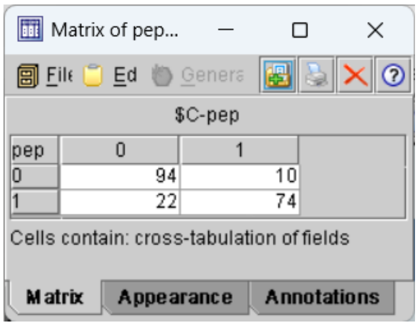


Figure 10: Minimum records per child branch (c) 10

### (b) Select the optimal decision tree

***Hand in: which of the three trees will you use to score the data in a holdout data list and why? 2-3 lines.***

### Confusion matrix for (a)

$$\begin{aligned} \text{precision} &= 0.7352941176470589 \\ \text{recall} &= 0.78125 \\ \text{F1} &= 0.7575757575757576 \end{aligned}$$

### Confusion matrix for (b)

$$\begin{aligned} \text{precision} &= 0.8780487804878049 \\ \text{recall} &= 0.75 \\ \text{F1} &= 0.8089887640449439 \end{aligned}$$

### Confusion matrix for (c)

$$\begin{aligned} \text{precision} &= 0.8809523809523809 \\ \text{recall} &= 0.7708333333333334 \\ \text{F1} &= 0.8222222222222222 \end{aligned}$$

The optimal decision tree is tree (c). Because the tree (c) F1 score is the highest, the tree (a) underfits the data, resulting in a lower metrics. The tree (b) and tree (c) gradually refine the branches of the tree, which helps to achieve consistent performance improvements without overfitting the data.

### (c) Predict the appendix data with the optimal decision tree

*Hand in: for the following data appendix, using the rules from the best decision tree, fill in the recommendation.*

Table (12 fields, 10 records) #1

| | age | sex | region | income | married | children | car | save_act | current_act | mortgage | $C-pep | $CC-pep |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 22 | 0 | 1 | 14000.... | 0 | 3 | 0 | 1 | 1 | 0 | 0 | 0.875 |
| 2 | 34 | 1 | 0 | 33000.... | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0.915 |
| 3 | 47 | 0 | 0 | 16700.... | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0.944 |
| 4 | 54 | 1 | 1 | 43400.... | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0.944 |
| 5 | 65 | 1 | 2 | 60000.... | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0.944 |
| 6 | 37 | 0 | 0 | 27700.... | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0.944 |
| 7 | 44 | 0 | 0 | 38784.... | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0.912 |
| 8 | 20 | 1 | 0 | 10200.... | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0.912 |
| 9 | 46 | 0 | 0 | 22000.... | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0.944 |
| 10 | 40 | 1 | 1 | 37400.... | 1 | 2 | 0 | 1 | 1 | 0 | 1 | 0.911 |

Table | Annotations

Figure 11: Prediction of the recommendation with the optimal decision tree (c)