# Homework 2

September 26, 2021

Jose Carlos Munoz

3.2c)

$$Gini_{Male} = 1 - (\frac{4}{10})^2 - (\frac{6}{10})^2 = 0.48 \quad (1)$$

$$Gini_{Female} = 1 - (\frac{4}{10})^2 - (\frac{6}{10})^2 = 0.48 \quad (2)$$

$$Gini_{Gender} = \frac{10}{20} * Gini_{Male} + \frac{10}{20} * Gini_{Female} = 0.48 \quad (3)$$

The Gini for Male is as showin in(1)
The Gini for Female is as showin in(2)
The Gini for gender is as showin in(3)

3.2d)

$$Gini_{Family} = 1 - (\frac{1}{4})^2 - (\frac{3}{4})^2 = 0.375 \quad (1)$$

$$Gini_{Sports} = 1 - (\frac{8}{8})^2 - (\frac{0}{8})^2 = 0.00 \quad (2)$$

$$Gini_{Luxury} = 1 - (\frac{1}{8})^2 - (\frac{7}{8})^2 = 0.21875 \quad (3)$$

$$Gini_{Cars} = \frac{4}{20} * Gini_{Family} + \frac{8}{20} * Gini_{Sports} + \frac{8}{20} * Gini_{Luxury} = 0.1625 \quad (4)$$

The Gini for Family is as showin in(2)
The Gini for Sports is as showin in(3)
The Gini for Luxury is as showin in(4)
The Gini for Cars is as showin in(**??**)

3.2e)

$$Gini_{Small} = 1 - (\frac{2}{5})^2 - (\frac{3}{5})^2 = .48 \quad (1)$$

$$Gini_{Medium} = 1 - (\frac{3}{7})^2 - (\frac{4}{7})^2 = \frac{24}{49} \quad (2)$$

$$Gini_{Large} = 1 - (\frac{3}{4})^2 - (\frac{1}{4})^2 = 0.5 \quad (3)$$

$$Gini_{Extra_Large} = 1 - (\frac{2}{4})^2 - (\frac{2}{4})^2 = 0.5 \quad (4)$$

$$Gini_{Shirt_size} = \frac{5}{20} * Gini_{Small} + \frac{7}{20} * Gini_{Medium} + \frac{4}{20} * Gini_{Large} + \frac{4}{20} * Gini_{Extra_Large} = 0.4914 \quad (5)$$

The Gini for Small is as showin in(1)
The Gini for Medium is as showin in(2)
The Gini for Large is as showin in(3)
The Gini for Extra Large is as showin in(4)
The Gini for Shirt Size is as showin in(5)

3.2f)
The Car type because it has the lowest Gini Index.

3.5a)

$$E_{orig} = -\frac{4}{10} * log(\frac{4}{10}) - \frac{6}{10} * log(\frac{6}{10}) = .9710 \quad (1)$$

The overall Entropy before the split is shown in (1)

$$\begin{aligned} E_T &= -\frac{4}{7} * log(\frac{4}{7}) - \frac{3}{7} * log(\frac{4}{7}) \\ E_F &= -\frac{3}{3} * log(\frac{3}{3}) - \frac{0}{3} * log(\frac{0}{0}) \\ \Delta E &= Eorig - \frac{7}{10} * E_T - \frac{3}{10} * E_F \\ &= 0.2813 \end{aligned} \quad (2)$$

The data gain from the splitting for A is show in (2)

$$\begin{aligned} E_T &= -\frac{3}{4} * log(\frac{3}{4}) - \frac{1}{4} * log(\frac{1}{4}) \\ E_F &= -\frac{1}{6} * log(\frac{1}{6}) - \frac{5}{6} * log(\frac{5}{6}) \\ \Delta E &= Eorig - \frac{4}{10} * E_T - \frac{6}{10} * E_F \\ &= 0.2565 \end{aligned} \quad (3)$$

The data gain from the splitting for B is show in (3)

3.7a)
P(A=1 |-)=2/5=0.4 , P(A=1 |+)=3/5=0.6 , P(A=0 |-)=3/5=0.6 , P(A=0 |+)=2/5=0.4
P(B=1 |-)=2/5=0.4 , P(B=1 |+)=1/5=0.2 , P(B=0 |-)=3/5=0.6 , P(B=0 |+)=4/5=0.8
P(C=1 |-)=4/4=1.0 , P(C=1 |+)=4/5=0.8 , P(C=0 |-)=0/4=0.0 , P(C=0 |+)=1/5=0.2

3.7b)
answer 1 3.7c)
answer 1
3.7a)
answer 1
3.7b)
answer 1 3.7c)