

Homework 2

September 26, 2021

Jose Carlos Munoz

3.2c)

$$\begin{aligned} Gini_{Male} &= 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{6}{10}\right)^2 \\ &= 0.48 \end{aligned} \tag{1}$$

$$\begin{aligned} Gini_{Female} &= 1 - \left(\frac{4}{10}\right)^2 - \left(\frac{6}{10}\right)^2 \\ &= 0.48 \end{aligned} \tag{2}$$

$$\begin{aligned} Gini_{Gender} &= \frac{10}{20} * Gini_{Male} + \frac{10}{20} * Gini_{Female} \\ &= 0.48 \end{aligned} \tag{3}$$

The Gini for Male is as shown in (1)

The Gini for Female is as shown in (2)

The Gini for gender is as shown in (3)

3.2d)

$$\begin{aligned} Gini_{Family} &= 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 \\ &= 0.375 \end{aligned} \tag{1}$$

$$\begin{aligned} Gini_{Sports} &= 1 - \left(\frac{8}{8}\right)^2 - \left(\frac{0}{8}\right)^2 \\ &= 0.00 \end{aligned} \tag{2}$$

$$\begin{aligned} Gini_{Luxury} &= 1 - \left(\frac{1}{8}\right)^2 - \left(\frac{7}{8}\right)^2 \\ &= 0.21875 \end{aligned} \tag{3}$$

$$\begin{aligned} Gini_{Cars} &= \frac{4}{20} * Gini_{Family} + \frac{8}{20} * Gini_{Sports} + \frac{8}{20} * Gini_{Luxury} \\ &= 0.1625 \end{aligned} \tag{4}$$

The Gini for Family is as shown in (2)

The Gini for Sports is as shown in (3)

The Gini for Luxury is as shown in (4)

The Gini for Cars is as shown in (??)

3.2e)

$$\begin{aligned} Gini_{Small} &= 1 - \left(\frac{2}{5}\right)^2 - \left(\frac{3}{5}\right)^2 \\ &= .48 \end{aligned} \tag{1}$$

$$\begin{aligned}
Gini_{Medium} &= 1 - \left(\frac{3}{7}\right)^2 - \left(\frac{4}{7}\right)^2 \\
&= \frac{24}{49}
\end{aligned} \tag{2}$$

$$\begin{aligned}
Gini_{Large} &= 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 \\
&= 0.5
\end{aligned} \tag{3}$$

$$\begin{aligned}
Gini_{ExtraLarge} &= 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 \\
&= 0.5
\end{aligned} \tag{4}$$

$$\begin{aligned}
Gini_{ShirtSize} &= \frac{5}{20} * Gini_{Small} + \frac{7}{20} * Gini_{Medium} + \frac{4}{20} * Gini_{Large} + \frac{4}{20} * Gini_{ExtraLarge} \\
&= 0.4914
\end{aligned} \tag{5}$$

The Gini for Small is as shown in (1)

The Gini for Medium is as shown in (2)

The Gini for Large is as shown in (3)

The Gini for Extra Large is as shown in (4)

The Gini for Shirt Size is as shown in (5)

3.2f)

The Car type because it has the lowest Gini Index.

3.5a)

$$\begin{aligned}
E_{orig} &= -\frac{4}{10} * \log\left(\frac{4}{10}\right) - \frac{6}{10} * \log\left(\frac{6}{10}\right) \\
&= .9710
\end{aligned} \tag{1}$$

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

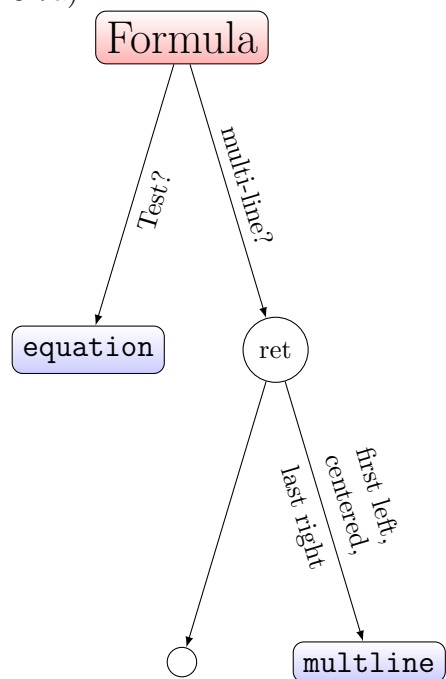
$$\begin{aligned}
E_T &= -\frac{4}{7} * \log\left(\frac{4}{7}\right) - \frac{3}{7} * \log\left(\frac{3}{7}\right) \\
E_F &= -\frac{3}{3} * \log\left(\frac{3}{3}\right) - \frac{0}{3} * \log\left(\frac{0}{3}\right) \\
\Delta E &= E_{orig} - \frac{7}{10} * E_T - \frac{3}{10} * E_F \\
&= 0.2813
\end{aligned} \tag{2}$$

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

$$\begin{aligned}
E_T &= -\frac{3}{4} * \log\left(\frac{3}{4}\right) - \frac{1}{4} * \log\left(\frac{1}{4}\right) \\
E_F &= -\frac{1}{6} * \log\left(\frac{1}{6}\right) - \frac{5}{6} * \log\left(\frac{5}{6}\right) \\
\Delta E &= E_{orig} - \frac{4}{10} * E_T - \frac{6}{10} * E_F \\
&= 0.2565
\end{aligned} \tag{3}$$

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

3.7a)



3.7b)

answer 1 3.7c)

answer 1

3.7a)

answer 1

3.7b)

answer 1 3.7c)