CSDS 440: Assignment 2

Shaochen (Henry) ZHONG, sxz517

Mingyang TIE, mxt497

Due on 09/18/2020, submitted early on 09/11/2020

Problem 8

With the new weight information, we have:

$$\begin{aligned} H(Y) &= -\frac{(3+9+3+9+3+9+27+81)}{256}\log_2(\frac{(3+9+3+9+3+9+27+81)}{256}) \\ &\quad -\frac{(1+3+9+27+9+27+9+27)}{256}\log_2(\frac{(1+3+9+27+9+27+9+27)}{256}) \\ &= -\frac{144}{256}\log_2(\frac{144}{256}) - \frac{112}{256}\log_2(\frac{112}{256}) \approx 0.9887 \end{aligned}$$

Now for each attribute:

$$
\begin{aligned}
IG(A_1) &= H(Y) - H(Y|A_1) \\
&= 0.9887 - [\frac{(3+9+9+27+9+27+27+81)}{256}\,(-\frac{(3+9+27+81)}{192}\log_2\frac{(3+9+27+81)}{192} \\
&\quad - \frac{(9+27+9+27)}{192}\log_2\frac{(3+9+27+81)}{192}) \\
&\quad + \frac{(1+3+3+9+3+9+9+27)}{256}\,(-\frac{(3+9+3+9)}{64}\log_2\frac{(3+9+3+9)}{64} \\
&\quad - \frac{(1+3+9+27)}{64}\log_2\frac{(1+3+9+27)}{64})] \\
&= 0.9887 - [\frac{192}{256}(-\frac{120}{192}\log_2(\frac{120}{192}) - \frac{72}{192}\log_2(\frac{72}{192})) + \frac{64}{256}(-\frac{24}{64}\log_2(\frac{24}{64}) - \frac{40}{64}\log_2\frac{40}{64})] \\
&\approx 0.0343
\end{aligned}
$$

Similarily:

$$
\begin{aligned}
IG(A_2) &= 0.9887 - [\frac{192}{256}(-\frac{120}{192}\log_2(\frac{120}{192}) - \frac{72}{192}\log_2(\frac{72}{192})) + \frac{64}{256}(-\frac{24}{64}\log_2(\frac{24}{64}) - \frac{40}{64}\log_2\frac{40}{64})] \\
&\approx 0.0343 \\
IG(A_3) &= 0.9887 - [\frac{192}{256}(-\frac{120}{192}\log_2(\frac{120}{192}) - \frac{72}{192}\log_2(\frac{72}{192})) + \frac{64}{256}(-\frac{24}{64}\log_2(\frac{24}{64}) - \frac{40}{64}\log_2\frac{40}{64})] \\
&\approx 0.0343 \\
IG(A_4) &= 0.9887 - [\frac{192}{256}(-\frac{108}{192}\log_2(\frac{108}{192}) - \frac{84}{192}\log_2(\frac{84}{192})) + \frac{64}{256}(-\frac{36}{64}\log_2(\frac{36}{64}) - \frac{28}{64}\log_2\frac{28}{64})] \\
&\approx 0
\end{aligned}
$$

Since we have $IG(A_1) \approx IG(A_2) \approx IG(A_3) \approx 0.0343$ and $IG(A_4) \approx 0$. Therefore the algorithm may first split on any attribute among A_1, A_2, or A_3 and form a decision tree.