CSDS 440: Assignment 2

Shaochen (Henry) ZHONG, `sxz517`

Mingyang TIE, `mxt497`

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

Problem 5

Since each Boolean attribute may have two outcomes, n Boolean attributes may lead to 2^n number of distinct examples (assuming every example is described by the same n Boolean attributes).

Problem 6

In pervious question we have showed if we put all examples, each with n Boolean attributes, into a table, such table will have 2^n rows.

Known that we may assign two possible class labels (1 or 0) for each example, and there are 2^n examples waiting for assignment. The question is equivalent to asking how many ways shall we fill a 2^n place array with 1s and 0s, and the answer would be 2^{2^n}.

Problem 7

For the entropy of table we have $H(Y) = -\frac{8}{16}\log_2(\frac{8}{16}) - \frac{8}{16}\log_2(\frac{8}{16}) = 1$

Known that $IG(A) = H(Y) - H(Y \mid A)$, and $H(Y \mid A) = P(A = T)H(Y \mid A = T) + P(A = F)H(Y \mid A = F)$. Since every attribute has the equal amount of Ts and Fs and also equal amount of 0 and 1 label output corresponding to their T and F groups, every attribute will have the following information gain:

$$IG(A_i) = H(Y) - H(Y \mid A_i) = H(Y) - [P(A_i = T)H(Y \mid A_i = T) + P(A_i = F)H(Y \mid A_i = F)]$$
$$= 1 - [\ \underbrace{\frac{8}{16}}_{P(A_i=T)}\ (\underbrace{-\frac{4}{8}\log_2(\frac{4}{8})}_{A_i=T\ and\ Y=1}\ \underbrace{-\frac{4}{8}\log_2(\frac{4}{8})}_{A_i=T\ and\ Y=0})\ +\ \underbrace{\frac{8}{16}}_{P(A_i=F)}\ (\underbrace{-\frac{4}{8}\log_2(\frac{4}{8})}_{A_i=F\ and\ Y=1}\ \underbrace{-\frac{4}{8}\log_2(\frac{4}{8})}_{A_i=F\ and\ Y=0}) = 0$$

Since every attribute has an IG of 0, and known that ID3 will stop if there is no infrmation gain. The algorithm will have no split at all.

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.

Problem 9

In *Problem 7*, every attribute has the equal amount of Ts and Fs and also equal amount of 0 and 1 labels corresponding to their T and F groups. Thus every attribute will have their $IG = 0$ and ID3 will stop.

However, in *Problem 8*, with the introduced weight, 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.