CSDS 440: Assignment 2

Shaochen (Henry) ZHONG, sxz517

Mingyang TIE, mxt497

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

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:

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

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.