Problem 2. Decision Trees [15]

(a) [3] The entropy of penny is: $H(P) = -\frac{2}{6}\log_2(\frac{2}{6}) - \frac{4}{6}\log_2(\frac{4}{6}) = 0.9183$

(b) [8] From (a) we know $H(P) = -\frac{2}{6}\log_2(\frac{2}{6}) - \frac{4}{6}\log_2(\frac{4}{6}) = 0.9183$

Because $H(P \mid Q) = \frac{5}{6}(-\frac{4}{5}\log_2(\frac{4}{5}) - \frac{1}{5}\log_2(\frac{1}{5})) + \frac{1}{6}(-\frac{1}{1}\log_2(\frac{1}{1}) - \frac{0}{1}\log_2(\frac{0}{1})) = 0.602$

the information gain between quarter and penny is:

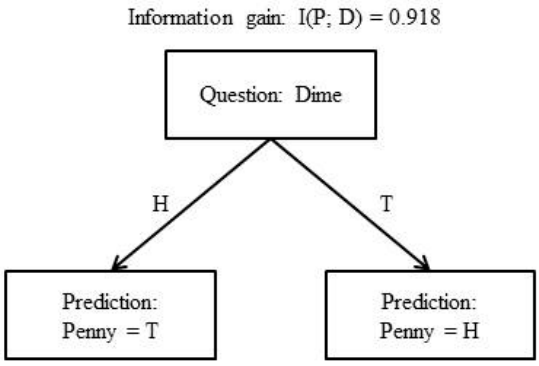
$I(P;Q) = H(P) - H(P \mid Q) = 0.317$

Similarly, $H(P \mid D) = \frac{4}{6}(-\frac{4}{4}\log_2(\frac{4}{4}) - \frac{0}{4}\log_2(\frac{0}{4})) + \frac{2}{6}(-\frac{0}{2}\log_2(\frac{0}{2}) - \frac{2}{2}\log_2(\frac{2}{2})) = 0$

and the information gain between dime and penny is:

$I(P;D) = H(P) - H(P \mid D) = 0.918$

(c) [2]



(d) [2] The training set accuracy is 100%. From the decision tree, we can see the dime can predict the penny "perfectly". This is an example of overfitting. If we didn't select the proper feature(s), or used a small training set, this can easily happen.