1. **Decision Trees and ID3**

   (a) The result of splitting on A:

   3:4

   1:2 2:2

   The associated entropy:

   $$-\left(\frac{1}{7}\ln\frac{1}{3}+\frac{2}{7}\ln\frac{2}{3}+\frac{2}{7}\ln\frac{2}{4}+\frac{2}{7}\ln\frac{2}{4}\right)\approx .669$$

   And for splitting on B:

   3:4

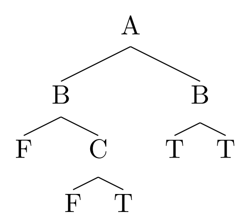
   2:3 1:1

   $$-\left(\frac{2}{7}\ln\frac{2}{5}+\frac{3}{7}\ln\frac{3}{5}+\frac{1}{7}\ln\frac{1}{2}+\frac{1}{7}\ln\frac{1}{2}\right)\approx .679$$

   So splitting on A provides a result with a slightly lower entropy, and hence slightly higher information gain.

   Splitting on A might be more useful because it provides a more even separation of the data into the true and false branches; splitting on B might be more useful because TODO.

   (b) In the following tree, going left indicates a True, and going right indicates a False.

   In the first stage, splitting on A generates the lowest remainder. Then considering the other three attributes, if we have A as True, then attributes B and C are tied for lowest remainder; if A is False, again B and C are tied for lowest remainder. We split on B in both cases.

   If A is True and B is True, there is only one case, in which the classification is False. Otherwise, there are two cases that are exactly dependent on C.

   If A is False and B is True, there is one case, in which the classification is True. Otherwise, there are two cases that cannot be separated, in which case we arbitrarily pick True.

   6 of 7 are correctly identified by the tree.

(c)

2. **ID3 with Pruning**

   (c)

   (d) i. TODO: how implementation makes use of instance weight in splitting decisions
   AdaBoost on ID3 would not work if splitting decisions were made by counting instances instead of summing weights, because AdaBoost is based on the idea that you update weights based on which data instances need better fitting.

   ii. Weighted entropy of $\{x_1, \ldots, x_n\}$ where $y_1 = T$, $y_i = F$ else, $w_1 = 0.5$, and other weights are $0.5/(n-1)$: by definition $entropy(D) = -\sum_{y \in Y} \frac{W_y}{W} \log_2 \frac{W_y}{W}$ and $W_y = \sum_i w_k(i) I(y_i = y)$; we have $W_T = 0.5$, $W_F = 0.5(n-1)^{-1}(n-1) = 0.5$, $W = 1$, and so

$$entropy(D) = -\left[\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right] = 1$$

   i.

   ii.

   iii.

   iv.

3.