1. **Decision Trees and ID3**

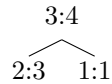   (a) The result of splitting on A:

   $$3:4$$

   $$1:2 \quad 2:2$$
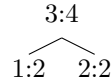
   The associated entropy:

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

   And for splitting on B:

   $$3:4$$

   $$2:3 \quad 1:1$$

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

   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)
   (c)

2. **ID3 with Pruning**

   (c)

   s 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.

   i. 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.