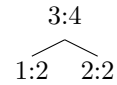


1. Decision Trees and ID3

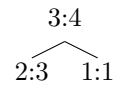
(a) The result of splitting on A :



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 :



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

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.

3.