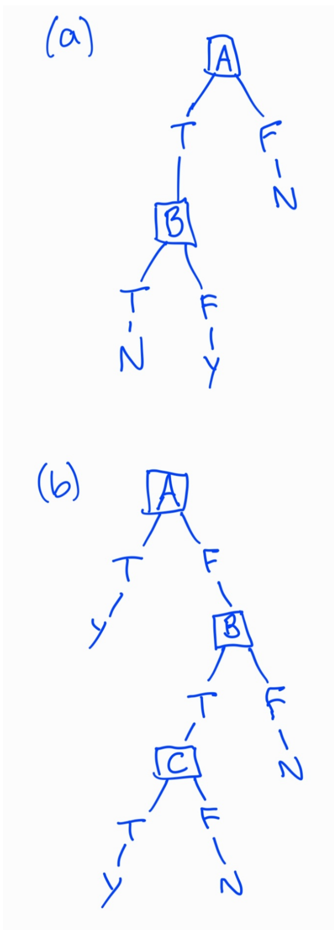
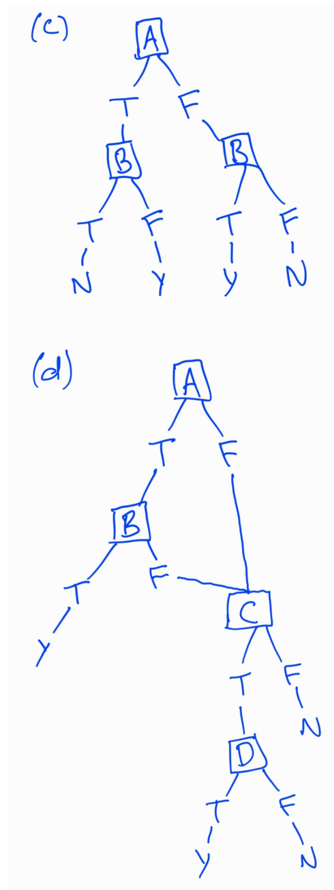
1:





2.1:

Sunny, overcast, rain.

2.2:

Sunny: D1, D2, D8, D9, D11
Overcast: D3, D7, D12, D13
Rain: D4, D5, D6, D10, D14

2.3:

Sunny → temperature:

$$E(S_{sunny}) = -\frac{2}{5}\log_2\frac{2}{5} - \frac{3}{5}\log_2\frac{3}{5} = 0.971$$

$$E(S_{hot}) = -0 - \frac{2}{2}\log_2\frac{2}{2} = 0$$

$$E(S_{mild}) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$E(S_{cool}) = -\frac{1}{1}\log_2\frac{1}{1} - 0 = 0$$

$$E(S_{sunny}|temperature) = \frac{2}{5}E(S_{hot}) + \frac{2}{5}E(S_{mild}) + \frac{1}{5}E(S_{cool}) = 0.4$$

$$G(S_{sunny}, temperature) = E(S_{sunny}) - E(S_{sunny}|temperature) = 0.571$$

Sunny → humidity:

$$E(S_{high}) = -0 - \frac{3}{3}\log_2\frac{3}{3} = 0$$

$$E(S_{normal}) = -\frac{2}{2}\log_2\frac{2}{2} - 0 = 0$$

$$E(S_{sunny}|temperature) = \frac{3}{5}E(S_{high}) + \frac{2}{5}E(S_{normal}) = 0$$

$$G(S_{sunny}, humidity) = E(S_{sunny}) - E(S_{sunny}|humidity) = 0.971$$

Sunny → wind:

$$E(S_{weak}) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$$

$$E(S_{strong}) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$E(S_{sunny}|temperature) = \frac{3}{5}E(S_{weak}) + \frac{2}{5}E(S_{strong}) = 0.9508$$

$$G(S_{sunny}, temperature) = E(S_{sunny}) - E(S_{sunny}|temperature) = 0.0202$$

Humidity has the highest gain so it should be the root for the sub-tree under sunny. No more subroots are needed under humidity and sunny because all outcomes of humidity have zero entropy.

Rain → temperature:

$$E(S_{rain}) = -\frac{3}{5}\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5} = 0.971$$

$$E(S_{hot}) = 0$$

$$E(S_{mild}) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918$$

$$E(S_{cool}) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$E(S_{rain}|temperature) = \frac{3}{5}E(S_{mild}) + \frac{2}{5}E(S_{cool}) = 0.951$$

$$G(S_{rain}, temperature) = E(S_{rain}) - E(S_{rain}|temperature) = 0.0202$$

Rain $\rightarrow$ humidity:

$$E(S_{high}) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$E(S_{normal}) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} = 0.918$$

$$E(S_{rain}|humidity) = \frac{2}{5}E(S_{high}) + \frac{3}{5}E(S_{normal}) = 0.9508$$

$$G(S_{rain}, humidity) = E(S_{rain}) - E(S_{rain}|humidity) = 0.0202$$

Rain $\rightarrow$ wind:

$$E(S_{strong}) = -0 - \frac{2}{2}\log_2\frac{2}{2} = 0$$

$$E(S_{weak}) = -\frac{3}{3}\log_2\frac{3}{3} - 0 = 0$$

$$E(S_{rain}|wind) = \frac{3}{5}E(S_{weak}) + \frac{2}{5}E(S_{strong}) = 0$$

$$G(S_{rain}, wind) = E(S_{rain}) - E(S_{rain}|wind) = 0.971$$

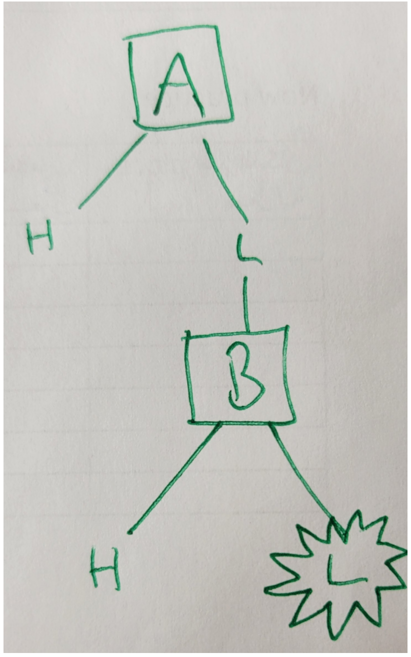
Wind has the highest gain so it should be the sub-root under rain. No more sub-roots are needed under wind and rain.

Overcast:

$$E(S_{overcast}) = -\frac{4}{4}\log_2\frac{4}{4} - 0 = 0$$

There is no sub-tree under overcast since it already has zero entropy.

3.1:



3.2:

If all leaf outcomes of the first decision tree are low-risk, then the second decision tree must be applied to every leaf of the first tree, so there would be $n = n_1 n_2$ leafs total.