

COMP 4983

Lab 7

Aaron Tansley

A00938978

October 28, 2022

* Samples in green denote a yes to play bocce

Forecast

Sunny	Overcast	Rain
1	3	4
2	7	5
8	12	6
9	13	10
11		14
# of Yes	2	4
# of No	3	0
Total	5	5

$$H_{\text{sunny}} = - \left(\frac{2}{5} \log_2 \frac{2}{5} + \frac{3}{5} \log_2 \frac{3}{5} \right)$$

$$= 0.971$$

$$H_{\text{overcast}} = - \left(\frac{4}{5} \log_2 \frac{4}{5} + \frac{1}{5} \log_2 \frac{1}{5} \right)$$

$$= 0$$

$$H_{\text{rain}} = - \left(\frac{1}{5} \log_2 \frac{1}{5} + \frac{4}{5} \log_2 \frac{4}{5} \right)$$

$$= 0.971$$

Temperature

Hot	Mild	Cold	
1	4	5	
2	8	6	
3	10	7	
13	11	9	
	12		
	14		
# of Yes	2	4	3
# of No	2	2	1
Total	4	6	4

$$H_{\text{Hot}} = - \left(\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right)$$

$$= 1$$

$$H_{\text{Mild}} = - \left(\frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6} \right)$$

$$= 0.918$$

$$H_{\text{Cold}} = - \left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right)$$

$$= 0.811$$

Humidity	
High	Normal
1	5
2	6
3	7
4	9
8	10
12	11
14	13
# of Yes	3
# of No	4
Total	7

Wind	
Strong	Weak
2	1
6	3
7	4
11	5
12	8
14	9
	10
	12
# of Yes	3
# of No	3
Total	6

$$H_{\text{High}} = -\left(\frac{3}{7} \log_2 \frac{3}{7} + \frac{4}{7} \log_2 \frac{4}{7}\right) \\ = 0.985$$

$$H_{\text{Strong}} = -\left(\frac{3}{6} \log_2 \frac{3}{6} + \frac{3}{6} \log_2 \frac{3}{6}\right) \\ = 1$$

$$H_{\text{Normal}} = -\left(\frac{6}{7} \log_2 \frac{6}{7} + \frac{1}{7} \log_2 \frac{1}{7}\right) \\ \approx 0.592$$

$$H_{\text{Weak}} = -\left(\frac{6}{8} \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8}\right) \\ = 0.811$$

$$\begin{aligned} \text{Total # of Yes} &= 9 \\ \text{Total # of No} &= 5 \end{aligned} \quad \left. \begin{array}{l} \text{Total} = 9+5 \\ = 14 \end{array} \right\}$$

$$H_{\text{root}} = -\left(\frac{9}{14} \log_2 \frac{9}{14} + \frac{5}{14} \log_2 \frac{5}{14}\right) \\ = 0.940$$

$$\text{Gain}(\text{root, Forecast}) = 0.940 - \left(\frac{5}{14}(0.971) + \frac{4}{14}(0) + \frac{5}{14}(0.971) \right)$$

$$= 0.246$$

$$\text{Gain}(\text{root, Temperature}) = 0.940 - \left(\frac{4}{14}(1) + \frac{6}{14}(0.918) + \frac{4}{14}(0.811) \right)$$

$$= 0.0291$$

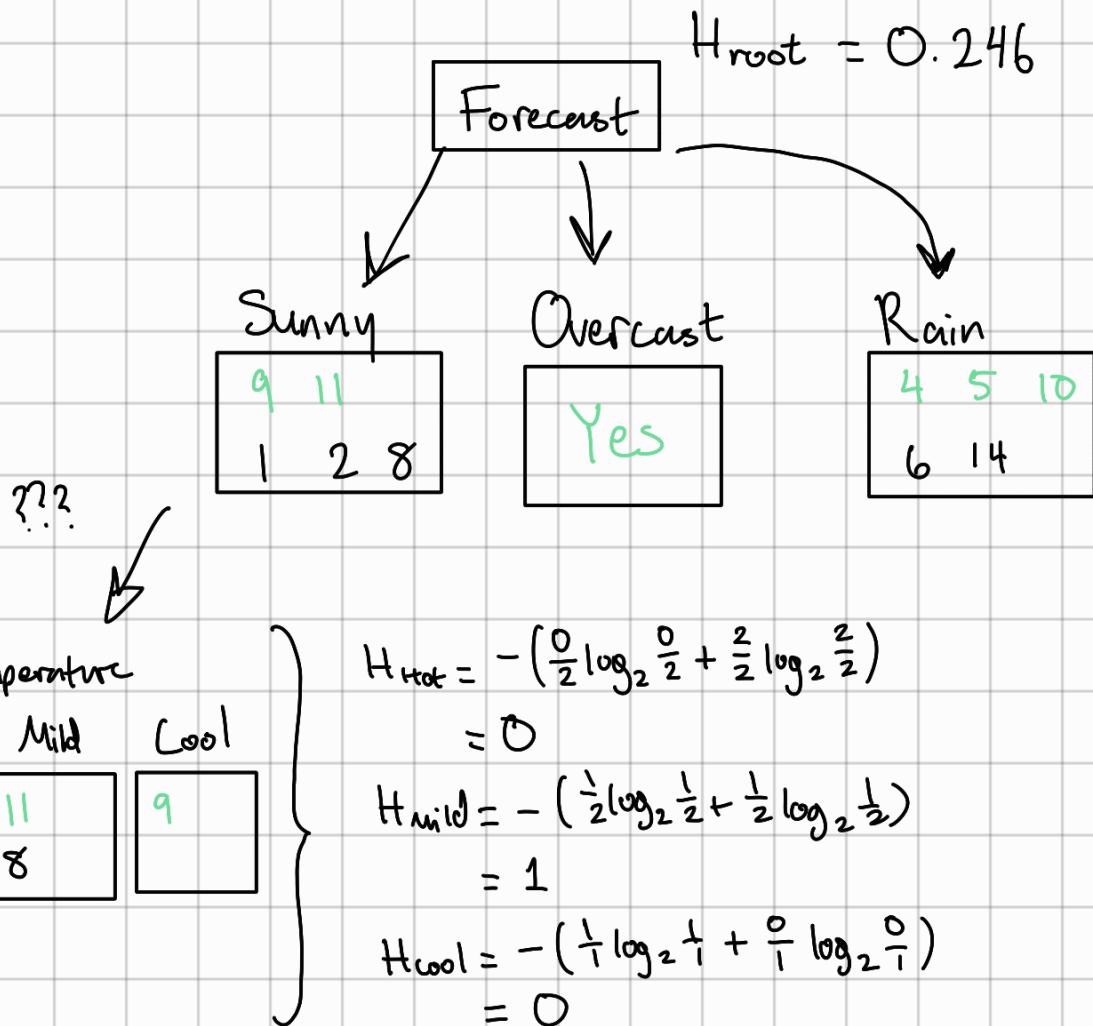
$$\text{Gain}(\text{root, Humidity}) = 0.940 - \left(\frac{7}{14}(0.985) + \frac{7}{14}(0.592) \right)$$

$$= 0.152$$

$$\text{Gain}(\text{root, Wind}) = 0.940 - \left(\frac{6}{14}(1) + \frac{8}{14}(0.811) \right)$$

$$= 0.0480$$

$\text{Gain}(\text{root, Forecast})$ has the largest value at 0.246, therefore it is selected as the root node.



	Humidity		
High	Normal		
1 2 8	9 11		

$$H_{\text{High}} = - \left(\frac{0}{3} \log_2 \frac{0}{3} + \frac{3}{3} \log_2 \frac{3}{3} \right) \\ = 0$$

$$H_{\text{Normal}} = - \left(\frac{2}{2} \log_2 \frac{2}{2} + \frac{0}{2} \log_2 \frac{0}{2} \right) \\ = 0$$

	Wind		
Weak	Strong		
9	11		
1 8	2		

$$H_{\text{Weak}} = - \left(\frac{1}{3} \log_2 \frac{1}{3} + \frac{2}{3} \log_2 \frac{2}{3} \right) \\ = 0.918$$

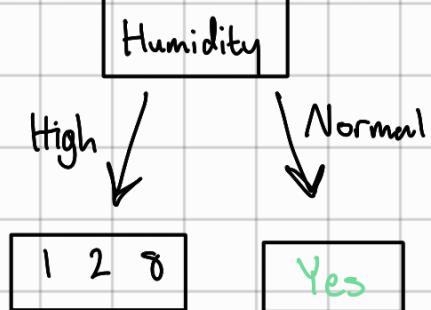
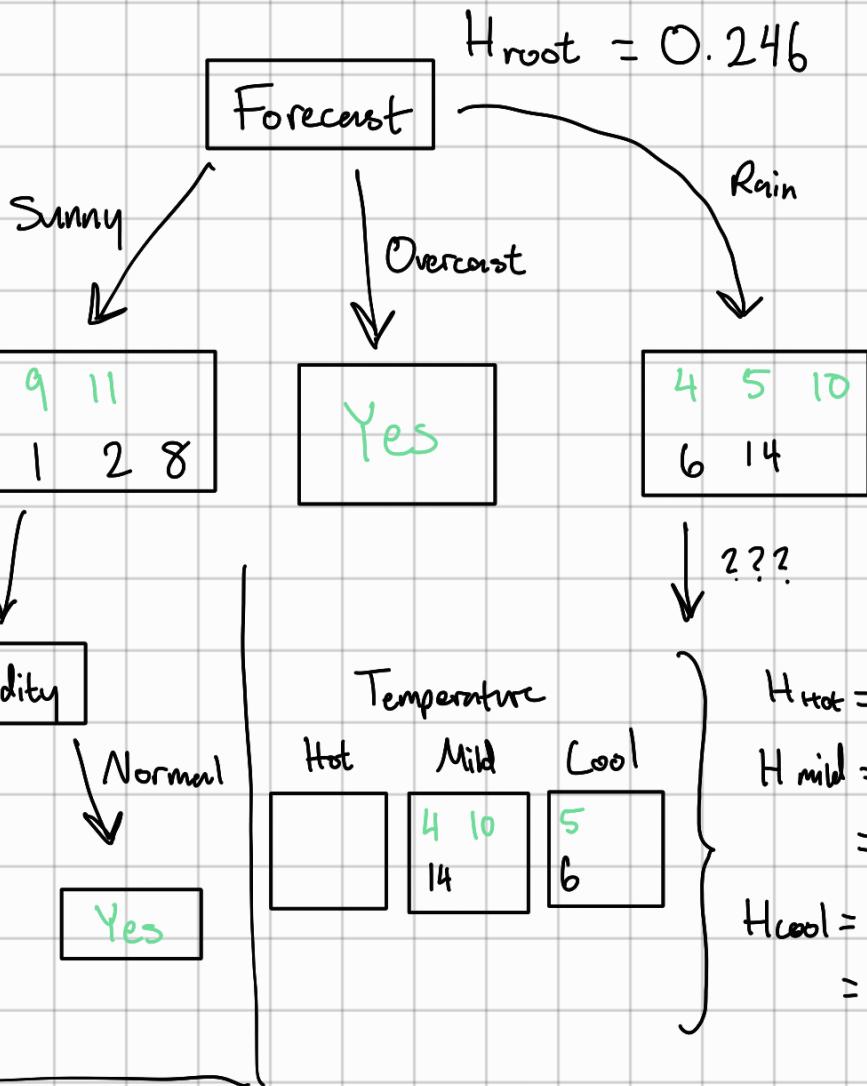
$$H_{\text{Strong}} = - \left(\frac{1}{2} \log_2 \frac{1}{2} + \frac{1}{2} \log_2 \frac{1}{2} \right) \\ = 1$$

$$\text{Gain}(\text{root}, \text{Temperature}) = 0.246 - \left(\frac{2}{5}(0) + \frac{2}{5}(1) + \frac{1}{5}(0) \right) \\ = -0.154$$

$$\text{Gain}(\text{root}, \text{Humidity}) = 0.246 - \left(\frac{3}{5}(0) + \frac{2}{5}(0) \right) \\ = 0.246$$

$$\text{Gain}(\text{root}, \text{Wind}) = 0.246 - \left(\frac{3}{5}(0.918) + \frac{2}{5}(1) \right) \\ = -0.7048$$

$\text{Gain}(\text{root}, \text{Humidity})$ has the largest gain of 0.246, therefore it is selected as the next root.



Temperature

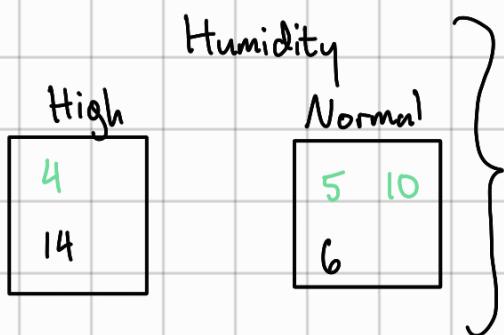


$$H_{root} = 0$$

$$H_{mild} = -\left(\frac{2}{3}\log_2 \frac{2}{3} + \frac{1}{3}\log_2 \frac{1}{3}\right) = 0.918$$

$$H_{cool} = -\left(\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{2}\log_2 \frac{1}{2}\right) = 1$$

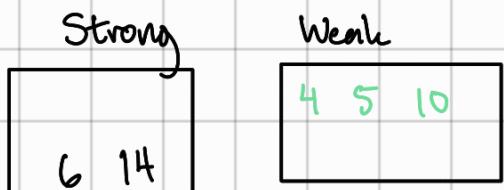
4, 5, 6, 10, 14



$$H_{high} = -\left(\frac{1}{2}\log_2 \frac{1}{2} + \frac{1}{2}\log_2 \frac{1}{2}\right) = 1$$

$$H_{normal} = -\left(\frac{2}{3}\log_2 \frac{2}{3} + \frac{1}{3}\log_2 \frac{1}{3}\right) = 0.918$$

Wind



$$H_{strong} = -\left(\frac{0}{2}\log_2 \frac{0}{2} + \frac{2}{2}\log_2 \frac{2}{2}\right) = 0$$

$$H_{weak} = -\left(\frac{3}{3}\log_2 \frac{3}{3} + \frac{0}{3}\log_2 \frac{0}{3}\right) = 0$$

$$\text{Gain}(\text{root, Temperature}) = 0.246 - \left(\frac{0}{5}(0) + \frac{3}{5}(0.918) + \frac{2}{5}(1) \right)$$

$$= -0.7048$$

$$\text{Gain}(\text{root, Humidity}) = 0.246 - \left(\frac{2}{5}(1) + \frac{3}{5}(0.918) \right)$$

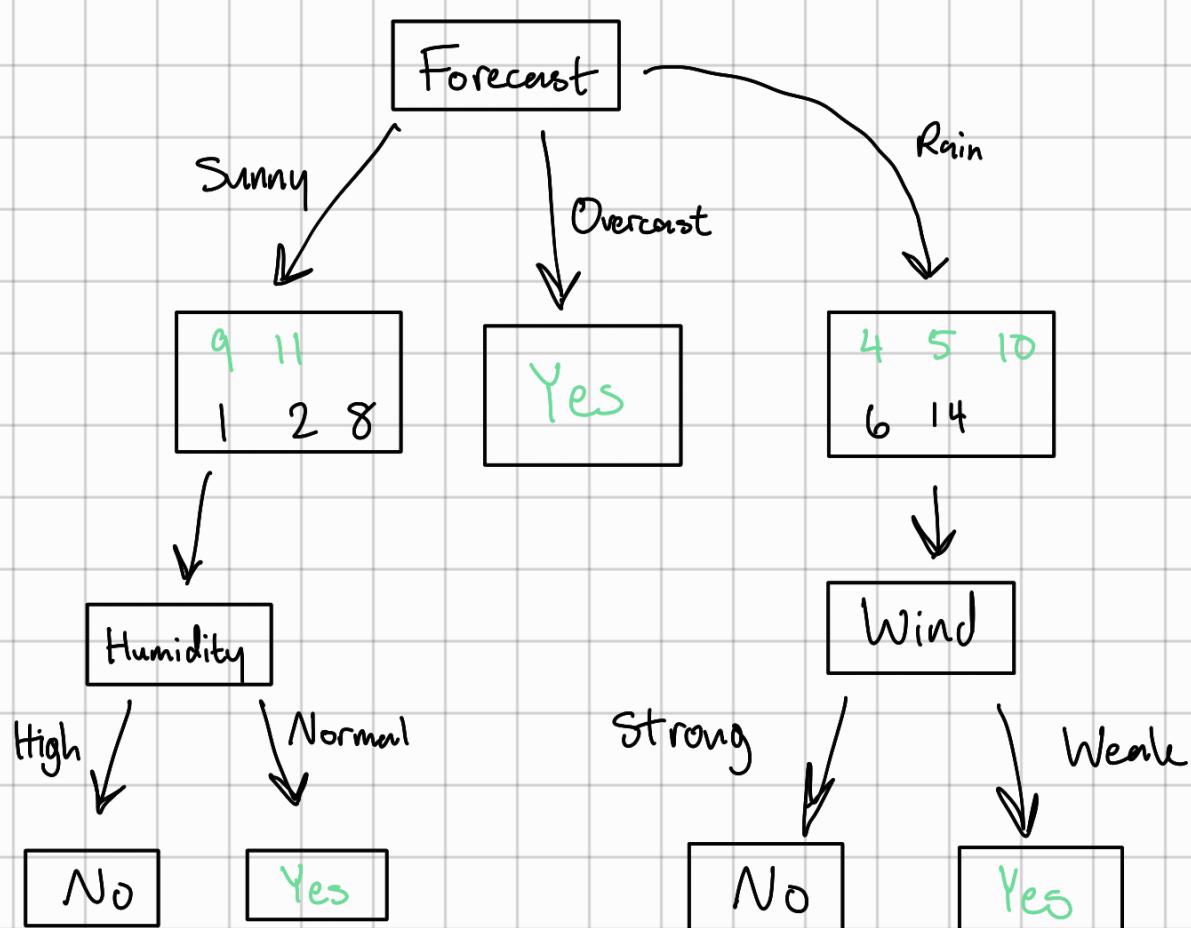
$$= -0.7048$$

$$\text{Gain}(\text{root, Wind}) = 0.246 - \left(\frac{2}{5}(0) + \frac{3}{5}(0) \right)$$

$$= 0.246$$

Since $\text{Gain}(\text{root, Wind})$ has the largest gain of 0.246, therefore it is chosen as the next root.

1) The decision tree may look something like:



2) It is NOT suitable to play bocce on a day that is (Sunny, Cool, High, Strong).