COEN 240 Machine Learning

Homework #5

Name: Alex Cherekdjian **Student ID:** 00001083236

Problem 1

1.1 Uncertainty before $= -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$

Uncertainty after:

$H(C|\text{Hardness}) = \frac{4}{8}H(C|\text{hard}) + \frac{4}{8}H(C|\text{soft})$

$= \frac{4}{8}H(\frac{3}{4},\frac{1}{4}) + \frac{4}{8}H(\frac{1}{4},\frac{3}{4})$

$= \frac{4}{8}\cdot(-\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}) + \frac{4}{8}(-\frac{1}{4}\log_2\frac{1}{4} - \frac{3}{4}\log_2\frac{3}{4})$

$= 0.5$

$H(C|\text{Grain}) = \frac{4}{8}H(C|\text{small}) + \frac{4}{8}H(C|\text{large})$

$= \frac{1}{2}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}) + \frac{1}{2}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2})$

$= \frac{1}{2}(1) + \frac{1}{2}(1) = 1$

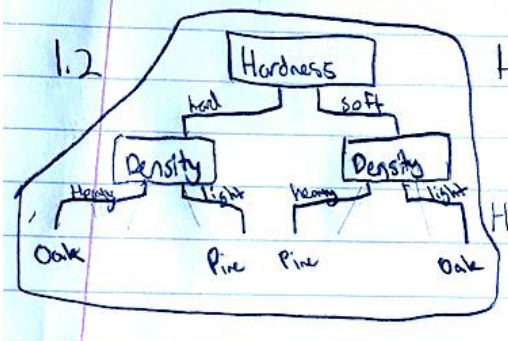
$H(C|\text{Density}) = \frac{6}{8}H(C|\text{Heavy}) + \frac{2}{8}H(C|\text{light})$

$= \frac{6}{8}(-\frac{3}{6}\log_2(\frac{3}{6}) - \frac{3}{6}\log_2\frac{3}{6}) + \frac{2}{8}(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2})$

$= \frac{6}{8}\cdot(1) + \frac{2}{8}(1) = 1$

Hardness will be chosen as the root.

1.2



$H(C|\text{Grain, hard}) = \frac{1}{2}H(\frac{1}{2},\frac{1}{2}) + \frac{1}{2}H(\frac{1}{2},\frac{1}{2})$ (small, large)

$= 1$

$H(C|\text{Grain, soft}) = \frac{1}{2}H(\frac{1}{2},\frac{1}{2}) + \frac{1}{2}H(\frac{1}{2},\frac{1}{2})$ (small, large)

$= 1$

$H(C|\text{Density, hard}) = \frac{3}{4}H(1,0) + \frac{1}{4}H(0,1)$ (Heavy, light)

$= 0$

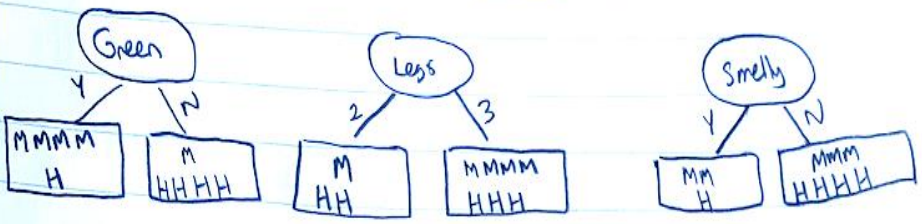
$H(C|\text{Density, soft}) = \frac{3}{4}H(0,1) + \frac{1}{4}H(1,0)$ (Heavy, light)

$= 0$

Problem 2

2a) $H(\text{Target Species}) = H(\tfrac{1}{2}, \tfrac{1}{2}) = \boxed{1 \text{ bit}}$

2b)



2c) $H(S \mid \text{Height}) = \frac{4}{10} H(\tfrac{1}{2}, \tfrac{1}{2}) + \frac{6}{10} H(\tfrac{3}{6}, \tfrac{3}{6})$ (Small, Tall)

$= \frac{4}{10}(1) + \frac{6}{10}(1)$

$= \boxed{1}$

2d) $H(S \mid \text{Green}) = \frac{5}{10} H(\tfrac{4}{5}, \tfrac{1}{5}) + \frac{5}{10} H(\tfrac{1}{5}, \tfrac{4}{5})$ (Yes, No)

$= \tfrac{1}{2}\left(-\tfrac{4}{5}\log_2 \tfrac{4}{5} - \tfrac{1}{5}\log_2 \tfrac{1}{5}\right) + \tfrac{1}{2}\left(-\tfrac{1}{5}\log_2 \tfrac{1}{5} - \tfrac{4}{5}\log_2 \tfrac{4}{5}\right)$

$= \boxed{0.722}$

2e) $H(S \mid \text{Legs}) = \frac{3}{10} H(\tfrac{1}{3}, \tfrac{2}{3}) + \frac{7}{10} H(\tfrac{4}{7}, \tfrac{3}{7})$ (2, 3)

$= \tfrac{3}{10}\left(-\tfrac{1}{3}\log_2 \tfrac{1}{3} - \tfrac{2}{3}\log_2 \tfrac{2}{3}\right) + \tfrac{7}{10}\left(-\tfrac{4}{7}\log_2 \tfrac{4}{7} - \tfrac{3}{7}\log_2 \tfrac{3}{7}\right)$

$= \boxed{0.9651}$

2f) $H(S \mid \text{Smelly}) = \frac{3}{10} H(\tfrac{2}{3}, \tfrac{1}{3}) + \frac{7}{10} H(\tfrac{3}{7}, \tfrac{4}{7})$ (Y)

$= \tfrac{3}{10}\left(-\tfrac{2}{3}\log_2 \tfrac{2}{3} - \tfrac{1}{3}\log_2 \tfrac{1}{3}\right) + \tfrac{7}{10}\left(-\tfrac{3}{7}\log_2 \tfrac{3}{7} - \tfrac{4}{7}\log_2 \tfrac{4}{7}\right)$

$= \boxed{0.9651}$

2g) **Green**, since it has the smallest entropy value & thus the most information gain of all the other attributes.

2h) $H(S|\text{Legs}, \text{yes}) = \frac{1}{5} H(\overset{2}{0}, \overset{''1''}{1}) + \frac{4}{5} H(\frac{4}{4}\overset{3}{0})$

$= 0$

$H(S|\text{Smelly}, \text{yes}) = \frac{2}{5} H(\overset{Y}{\frac{1}{2}}, \frac{1}{2}) + \frac{3}{5} H(\overset{N}{\frac{3}{3}}, \frac{0}{3})$

$= \frac{2}{5} = 0.4$

$H(S|\text{Height}, \text{yes}) = \frac{2}{5} H(\overset{\text{Small}}{\frac{2}{2}}, \frac{0}{2}) + \frac{3}{5} H(\overset{\text{tall}}{\frac{2}{3}}, \frac{1}{3})$

$= \frac{3}{5}(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3})$

$= 0.551$

choose "Legs" for "Green = yes"

$H(S|\text{Legs}, \text{No}) = \frac{2}{5} H(\overset{2}{\frac{1}{2}}, \frac{1}{2}) + \frac{3}{5} H(\overset{3}{0}, 1)$

$= \frac{2}{5} = 0.4$

$H(S|\text{Smelly}, \text{No}) = \frac{1}{5} H(\overset{Y}{1}, 0) + \frac{4}{5} H(\overset{N}{0}, \frac{4}{4})$

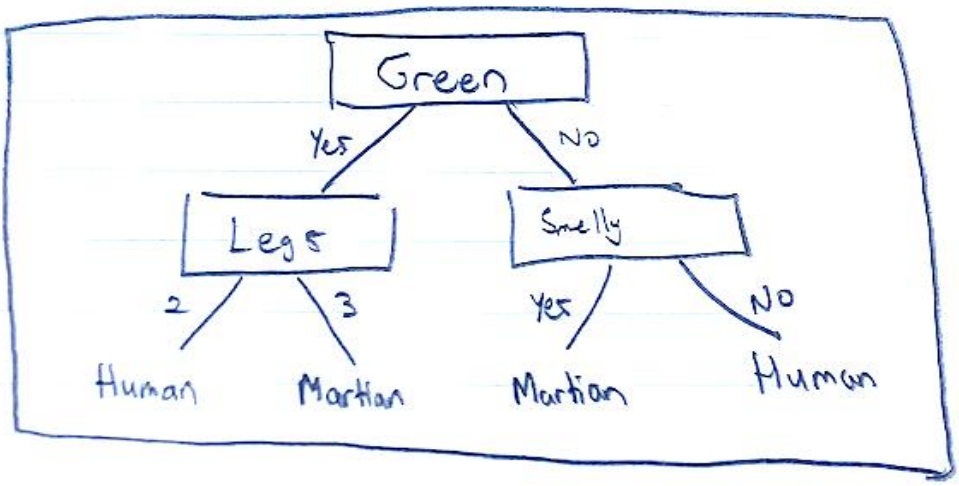
$= 0$

$H(S|\text{Height}, \text{No}) = \frac{2}{5} H(\overset{\text{Small}}{\frac{0}{2}}, \frac{2}{2}) + \frac{3}{5} H(\overset{\text{tall}}{\frac{1}{3}}, \frac{2}{3})$

$= \frac{3}{5}(-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3})$

$= 0.551$

choose "Smelly" for "Green = no"



Attachment

None