Jon Rippe
CSCE A415
HW #1.1

Problem 1: Decision Trees

Size	Orbit	Habitable	Count
big	near	yes	20
big	far	yes	170
small	near	yes	139
small	far	yes	45
big	near	no	130
big	far	no	30
small	near	no	11
small	far	no	255

Determine Overall Entropy:

$$E = H(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

Overall:

- $\#S = 20 + 170 + 139 + 45 + 130 + 30 + 11 + 255 = 800$
- $P(\text{yes}) = p_+ = \frac{20+170+139+45}{800} = \frac{374}{800} = .4675$
- $P(\text{no}) = p_- = 1 - .4675 = .5325$
- $E(\text{overall}) = -.4675 \log_2(.4675) - .5325 \log_2(.5325) = .9970$

Find the best Gain:

Size:

- Big
  - $\#S = 20 + 170 + 130 + 30 = 350$
  - $p_+ = \frac{20+170}{350} = \frac{190}{350} = .5429$
  - $p_- = 1 - .5429 = .4571$
  - $E(\text{big}) = -.5429 \log_2(.5429) - .4571 \log_2(.4571) = .9947$
- Small
  - $\#S = 139 + 45 + 11 + 255 = 450$
  - $p_+ = \frac{139+45}{450} = \frac{184}{450} = .4089$
  - $p_- = 1 - .4089 = .5911$
  - $E(\text{small}) = -.4089 \log_2(.4089) - .5911 \log_2(.5911) = .9759$
- $Gain(\text{size}) = .9970 - \left(\frac{350}{800}\right).9947 - \left(\frac{450}{800}\right).9759 = .0129$

Orbit:

- Near
  - $\#S = 20 + 139 + 130 + 11 = 300$
  - $p_+ = \frac{20+139}{300} = \frac{159}{300} = .5300$
  - $p_- = 1 - .5300 = .4700$
  - $E(\text{near}) = -.5300 \log_2(.5300) - .4700 \log_2(.4700) = .9974$
- Far
  - $\#S = 170 + 45 + 30 + 255 = 500$
  - $p_+ = \frac{170+45}{500} = \frac{215}{500} = .4300$
  - $p_- = 1 - .5700 = .5700$
  - $E(\text{far}) = -.4300 \log_2(.4300) - .5700 \log_2(.5700) = .9858$
- $Gain(\text{orbit}) = .9970 - \left(\frac{300}{800}\right).9974 - \left(\frac{500}{800}\right).9858 = .0069$

Start the Decision Tree:

There are only two attributes, *size* and *orbit*. The tree will start with *size* because it shows the most gain in the calculations.

Calculate Orbit Probabilities

Calculating the probability of Y given both attributes for all combinations will give the final Yes/No leaves.

$$P(Y|\text{big} \wedge \text{near}) = \frac{20}{150} = .1333$$
$$P(\bar{Y}|\text{big} \wedge \text{near}) = 1 - .1333 = .8667$$
$$P(Y|\text{big} \wedge \text{far}) = \frac{170}{200} = .8500$$
$$P(\bar{Y}|\text{big} \wedge \text{far}) = 1 - .8500 = .1500$$
$$P(Y|\text{small} \wedge \text{near}) = \frac{139}{150} = .9267$$
$$P(\bar{Y}|\text{small} \wedge \text{near}) = 1 - .9267 = .0733$$
$$P(Y|\text{small} \wedge \text{far}) = \frac{45}{300} = .1500$$
$$P(\bar{Y}|\text{small} \wedge \text{far}) = 1 - .1500 = .8500$$