Computer Science Assignment Question Two

Michael Alizzi
15 April 2025

---

a) Manually construct a Decision Tree for the following set of examples. What class is assigned to the instance of D15: {Sunny, Hot, High, Weak}?

$$\text{Entropy}_S = -p_+ \cdot \log_2(p_+) - p_- \cdot \log_2(p_-)$$

$$H(Y \mid X = x_j) = -\sum_{i=1}^{n} P(y_i \mid X = x_j) \log_2 P(y_i \mid X = x_j)$$

$$\mathrm{I(X, Y) = H(Y) - H(Y|X)}$$

$$\text{Entropy}_S = -\frac{9}{14} \cdot \log_2 \frac{9}{14} - \frac{5}{14} \cdot \log_2 \frac{5}{14} = 0.940$$

$$\text{Entropy}_{\text{Sunny}} = -\frac{2}{5} \cdot \log_2 \left(\frac{2}{5}\right) - \frac{3}{5} \cdot \log_2 \left(\frac{3}{5}\right) = 0.971$$

$$\text{Entropy}_{\text{Overcast}} = -\frac{4}{4} \cdot \log_2 \left(\frac{4}{4}\right) - \frac{0}{4} \cdot \log_2 \left(\frac{0}{4}\right) = 0$$

$$\text{Entropy}_{\text{Rain}} = -\frac{3}{5} \cdot \log_2 \left(\frac{3}{5}\right) - \frac{2}{5} \cdot \log_2 \left(\frac{2}{5}\right) = 0.971$$

$$\text{Entropy}_{\text{Outlook}} = 0.971 \cdot \frac{5}{14} + 0 \cdot \frac{4}{14} + 0.971 \cdot \frac{5}{14} = 0.694$$

$$\text{Information Gain}_{\text{Outlook}} = 0.940 - 0.694 = 0.246$$

$$\text{Entropy}_{\text{Hot}} = -\frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) - \frac{2}{4} \cdot \log_2 \left(\frac{2}{4}\right) = 1$$

$$\text{Entropy}_{\text{Mild}} = -\frac{4}{6} \cdot \log_2 \left(\frac{4}{6}\right) - \frac{2}{6} \cdot \log_2 \left(\frac{2}{6}\right) = 0.918$$

$$\text{Entropy}_{\text{Cool}} = -\frac{3}{4} \cdot \log_2 \left(\frac{3}{4}\right) - \frac{1}{4} \cdot \log_2 \left(\frac{1}{4}\right) = 0.811$$

$$\text{Entropy}_{\text{Tempuraturе}} = 1 \cdot \frac{4}{14} + 0.918 \cdot \frac{6}{14} + 0.811 \cdot \frac{4}{14} = 0.911$$

$$\text{Information Gain}_{\text{Temperature}} = 0.940 - 0.911 = 0.029$$

$$\text{Entropy}_{\text{High}} = -\frac{3}{7} \cdot \log_2 \left(\frac{3}{7}\right) - \frac{4}{7} \cdot \log_2 \left(\frac{4}{7}\right) = 0.985$$

$$\text{Entropy}_{\text{Normal}} = -\frac{6}{7} \cdot \log_2 \left(\frac{6}{7}\right) - \frac{1}{7} \cdot \log_2 \left(\frac{1}{7}\right) = 0.592$$

$$\text{Entropy}_{\text{Humidity}} = 0.985 \cdot \frac{7}{14} + 0.592 \cdot \frac{7}{14} = 0.788$$

$$\text{Information Gain}_{\text{Humidity}} = 0.940 - 0.788 = 0.152$$

$$\text{Entropy}_{\text{Weak}} = -\frac{6}{8} \cdot \log_2 \left(\frac{6}{8}\right) - \frac{2}{8} \cdot \log_2 \left(\frac{2}{8}\right) = 0.811$$

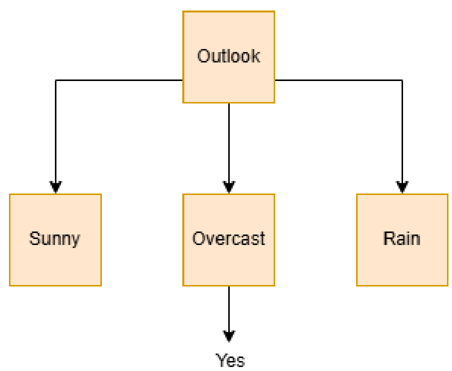
$$\text{Entropy}_{\text{Strong}} = -\frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right) - \frac{3}{6} \cdot \log_2\left(\frac{3}{6}\right) = 1$$

$$\text{Entropy}_{\text{Wind}} = 0.811 \cdot \frac{8}{14} + 1 \cdot \frac{6}{14} = 0.892$$

$$\text{Information Gain}_{\text{Wind}} = 0.940 - 0.892 = 0.048$$

Outlook: $0.246 > (0.029, 0.151, 0.048)$

$$p_{+}(\text{Overcast}) = \frac{4}{4}$$



$$\text{Entropy}_{\text{Sunny}} = -\frac{2}{5} \cdot \log_2 \frac{2}{5} - \frac{3}{5} \cdot \log_2 \frac{3}{5} = 0.971$$

$$\text{SunnyEntropy}_{\text{Hot}} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

$$\text{SunnyEntropy}_{\text{Mild}} = -\frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log_2\left(\frac{1}{2}\right) = 1$$

$$\text{SunnyEntropy}_{\text{Cool}} = -\frac{1}{1} \cdot \log_2\left(\frac{1}{1}\right) - \frac{0}{1} \cdot \log_2\left(\frac{0}{1}\right) = 0$$

$$\text{SunnyEntropy}_{\text{Tempurature}} = 0 \cdot \frac{2}{5} + 1 \cdot \frac{2}{5} + 0 \cdot \frac{1}{5} = 0.4$$

$$\text{Information Gain}_{\text{Temperature}} = 0.971 - 0.4 = 0.571$$

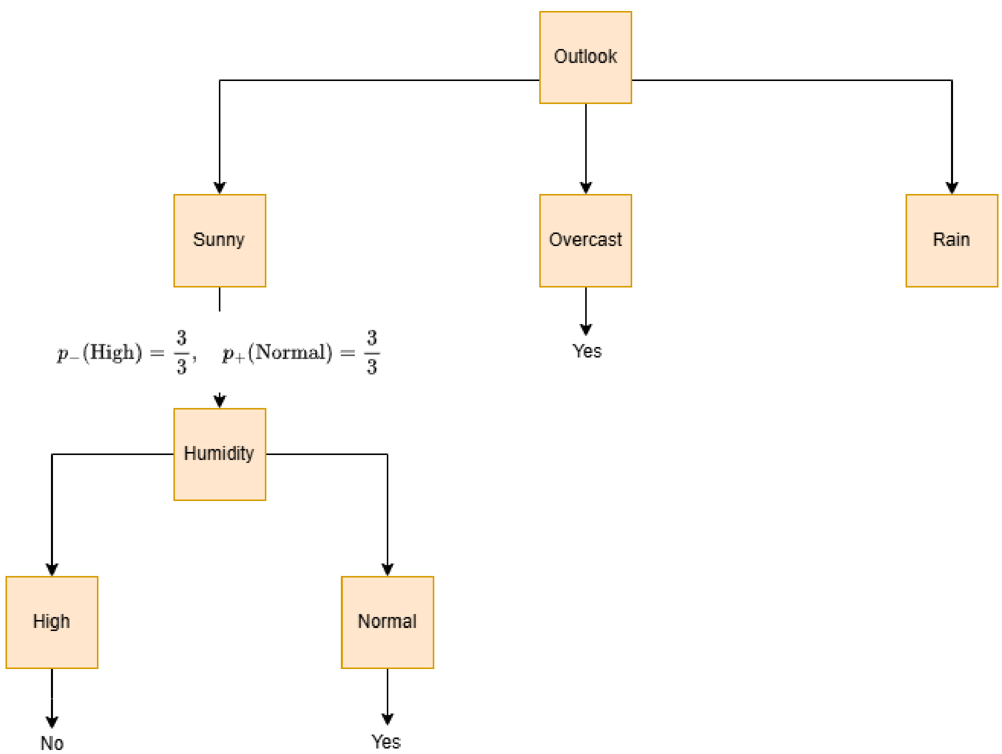
$$\text{SunnyEntropy}_{\text{High}} = -\frac{0}{3} \cdot \log_2\left(\frac{0}{3}\right) - \frac{3}{3} \cdot \log_2\left(\frac{3}{3}\right) = 0$$

$$\text{SunnyEntropy}_{\text{Normal}} = -\frac{2}{2} \cdot \log_2\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log_2\left(\frac{0}{2}\right) = 0$$

$$\text{SunnyEntropy}_{\text{Humidity}} = 0 \cdot \frac{3}{5} + 0 \cdot \frac{2}{5} = 0$$

$$\text{Information Gain}_{\text{Humidity}} = 0.971 - 0 = 0.971$$

$$\text{Information Gain}_{\text{Humidity}} = \max(\text{Information Gain})$$



$$\text{Entropy}_{\text{Rain}} = -\frac{3}{5}\cdot\log_3\frac{2}{5} - \frac{2}{5}\cdot\log_2\frac{2}{5} = 0.971$$

$$\text{RainEntropy}_{\text{Hot}} = 0$$

$$\text{RainEntropy}_{\text{Mild}} = -\frac{2}{3}\cdot\log_2\left(\frac{2}{3}\right) - \frac{1}{3}\cdot\log_2\left(\frac{1}{3}\right) = 0.918$$

$$\text{RainEntropy}_{\text{Cool}} = -\frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\cdot\log_2\left(\frac{1}{2}\right) = 1$$

$$\text{RainEntropy}_{\text{Tempurature}} = 0.918\cdot\frac{3}{5} + 1\cdot\frac{2}{5} = 0.951$$

$$\text{Information Gain}_{\text{Tempurature}} = 0.971 - 0.951 = 0.02$$

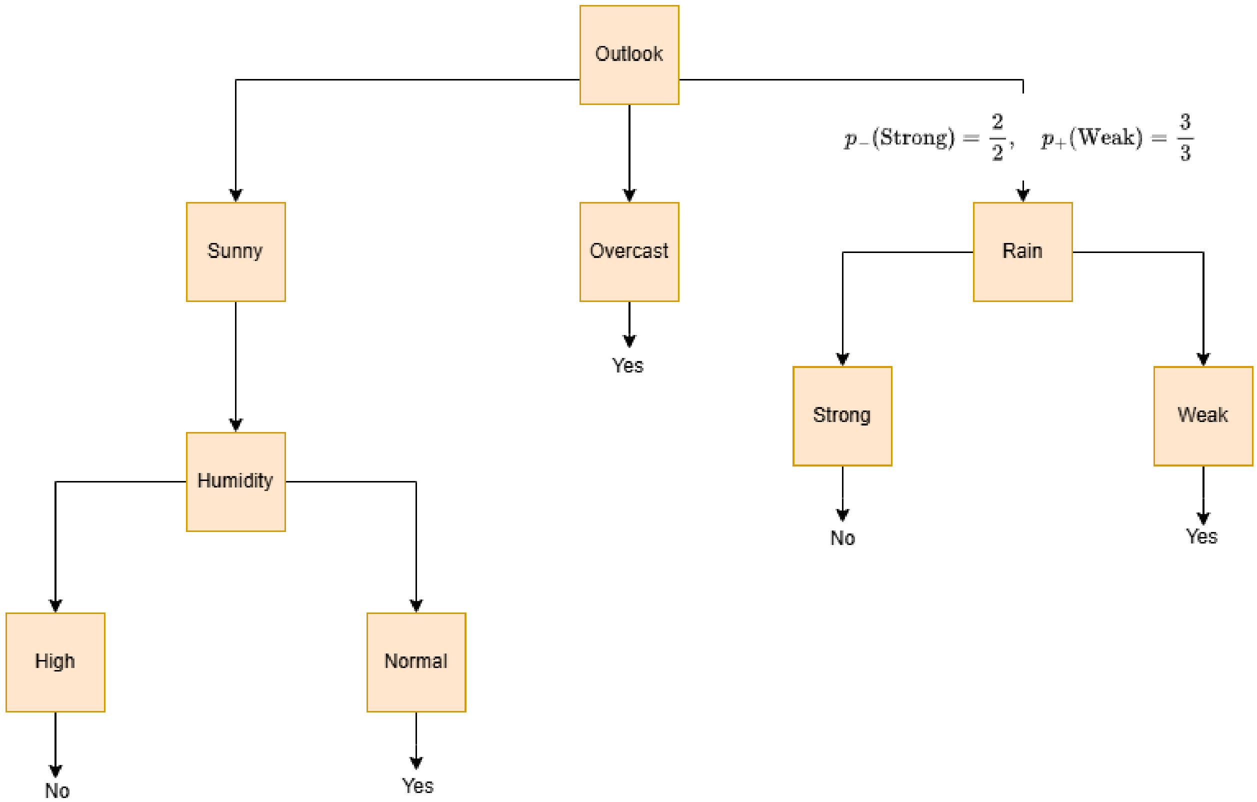
$$\text{RainEntropy}_{\text{Weak}} = -\frac{3}{3}\cdot\log_2\left(\frac{3}{3}\right) - \frac{0}{3}\cdot\log_2\left(\frac{0}{3}\right) = 0$$

$$\text{RainEntropy}_{\text{Strong}} = -\frac{0}{2}\cdot\log_2\left(\frac{0}{2}\right) - \frac{2}{2}\cdot\log_2\left(\frac{2}{2}\right) = 0$$

$$\text{RainEntropy}_{\text{Wind}} = 0\cdot\frac{3}{5} + 0\cdot\frac{2}{5} = 0$$

$$\text{Information Gain}_{\text{Wind}} = 0.971 - 0 = 0.971$$

$$\text{Information Gain}_{\text{Wind}} = \max(\text{Information Gain})$$



Following the tree, D15 would be classified as no (or not playing tennis).

b) Consider a Naive Bayes classifier for the same set of examples. What class is assigned to the instance of D15: {Sunny, Hot, High, Weak}?

$$\text{P(Play | Outlook, Temperature, Humidity, Wind)} =$$

$$\frac{\text{P(Outlook, Temperature, Humidity, Wind | Play)} \cdot \text{P(Play)}}{\text{P(Outlook, Temperature, Humidity, Wind)}}$$

P(Outlook, Temperature, Humidity, Wind $|\text{Play}_{\text{Yes}}) \cdot P(\text{Play}_{\text{Yes}}) =$

$P(\text{Sunny} \mid \text{Play}_{\text{Yes}}) \cdot P(\text{Hot} \mid \text{Play}_{\text{Yes}}) \cdot P(\text{High} \mid \text{Play}_{\text{Yes}}) \cdot P(\text{Weak} \mid \text{Play}_{\text{Yes}}) \cdot P(\text{Play}_{\text{Yes}}) =$

$\frac{2}{9} \cdot \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{6}{9} \cdot \frac{9}{14} = 0.007$

P(Outlook, Temperature, Humidity, Wind $|\text{Play}_{\text{No}}) \cdot P(\text{Play}_{\text{No}}) =$

$P(\text{Sunny} \mid \text{Play}_{\text{No}}) \cdot P(\text{Hot} \mid \text{Play}_{\text{No}}) \cdot P(\text{High} \mid \text{Play}_{\text{No}}) \cdot P(\text{Weak} \mid \text{Play}_{\text{No}}) \cdot P(\text{Play}_{\text{No}}) =$

$\frac{3}{5} \cdot \frac{2}{5} \cdot \frac{4}{5} \cdot \frac{2}{5} \cdot \frac{5}{14} = 0.027$

P(Outlook, Temperature, Humidity, Wind) = constant

Therefore:

$0.027 > 0.007$

The class assigned for D15 is no (or not playing tennis).