Decision Tree for Play Tennis

Md. Abdullah Al KafiLecturer, Department Of CSEDaffodil International University

September 12, 2024

1 Problem Overview

We have the following dataset for determining whether to play tennis based on weather conditions and temperature:

Sample	**Weather**	**Temperature**	**Play Tennis**
1	Sunny	Hot	No
2	Sunny	Hot	No
3	Overcast	Hot	Yes
4	Rainy	Mild	Yes
5	Rainy	Cool	Yes
6	Rainy	Cool	No
7	Overcast	Cool	Yes
8	Sunny	Mild	No
9	Sunny	Cool	Yes
10	Rainy	Mild	Yes

Table 1: Dataset

2 Step 1: Calculate Entropy for the Target Variable

We calculate the entropy for the target variable *Play Tennis*. The distribution is:

$$\text{Yes} = 6, \quad \text{No} = 4$$

The formula for entropy is:

$$\text{Entropy}(S) = -\left(\frac{6}{10}\log_2\frac{6}{10} + \frac{4}{10}\log_2\frac{4}{10}\right)$$

$$\text{Entropy}(S) = -\left(0.6\log_2 0.6 + 0.4\log_2 0.4\right) = 0.971$$

3 Step 2: Calculate Information Gain for Features

3.1 Information Gain for Weather

We split the dataset based on weather conditions: Sunny, Overcast, and Rainy.

For Sunny:

- Yes = 1, No = 3

$$\text{Entropy(Sunny)} = -\left(\frac{1}{4}\log_2\frac{1}{4} + \frac{3}{4}\log_2\frac{3}{4}\right) = 0.811$$

For Overcast:

- Yes = 2, No = 0 (Pure subset)

$$\text{Entropy(Overcast)} = 0$$

For Rainy:

- Yes = 3, No = 1

$$\text{Entropy(Rainy)} = -\left(\frac{3}{4}\log_2\frac{3}{4} + \frac{1}{4}\log_2\frac{1}{4}\right) = 0.811$$

Weighted Average Entropy for Weather:

$$\text{Entropy(Weather)} = \frac{4}{10} \times 0.811 + \frac{2}{10} \times 0 + \frac{4}{10} \times 0.811 = 0.648$$

Information Gain for Weather:

$$\text{Gain(Weather)} = 0.971 - 0.648 = 0.323$$

3.2 Information Gain for Temperature

We now calculate the information gain for Temperature by splitting the dataset into Hot, Mild, and Cool.

For Hot:

- Yes = 1, No = 2

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

For Mild:

- Yes = 2, No = 1

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

For Cool:

- Yes = 3, No = 1

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

Weighted Average Entropy for Temperature:

$$\text{Entropy(Temperature)} = \frac{3}{10} \times 0.918 + \frac{3}{10} \times 0.918 + \frac{4}{10} \times 0.811 = 0.88$$

Information Gain for Temperature:

$$\text{Gain(Temperature)} = 0.971 - 0.88 = 0.091$$

4 Step 3: Build the Decision Tree

Since **Weather** gives the highest information gain, we split first on Weather.

4.1 If Weather = Overcast:

$$\text{Play Tennis} = \text{Yes} \quad \text{(Pure Subset)}$$

4.2 If Weather = Sunny:

We now split on Temperature for Sunny weather:

- If Temperature = Hot, Play Tennis = No
- If Temperature = Mild, Play Tennis = No
- If Temperature = Cool, Play Tennis = Yes

4.3 If Weather = Rainy:

We now split on Temperature for Rainy weather:

- If Temperature = Mild, Play Tennis = Yes
- If Temperature = Cool, Play Tennis = Yes (Majority)

5 Final Decision Tree

The final decision tree is as follows:

- If Weather = Overcast, Play Tennis = Yes
- If Weather = Sunny:
  - If Temperature = Hot, Play Tennis = No
  - If Temperature = Mild, Play Tennis = No
  - If Temperature = Cool, Play Tennis = Yes
- If Weather = Rainy:
  - If Temperature = Mild, Play Tennis = Yes
  - If Temperature = Cool, Play Tennis = Yes