Decision Tree for Buys Computer Problem

Your Name

1 Dataset

The dataset contains information about customers, and we aim to predict whether they buy a computer based on attributes like Age, Income, Student status, and Credit Rating.

RID	**Age**	**Income**	**Student**	**Credit Rating**	**Buys Computer?**
1	Youth	High	No	Fair	No
2	Youth	High	No	Excellent	No
3	Middle Aged	High	No	Fair	Yes
4	Senior	Medium	No	Fair	Yes
5	Senior	Low	Yes	Fair	Yes
6	Senior	Low	Yes	Excellent	No
7	Middle Aged	Low	Yes	Excellent	Yes
8	Youth	Medium	No	Fair	No
9	Youth	Low	Yes	Fair	Yes
10	Senior	Medium	Yes	Fair	Yes
11	Youth	Medium	Yes	Excellent	Yes
12	Middle Aged	Medium	No	Excellent	Yes
13	Middle Aged	High	Yes	Fair	Yes
14	Senior	Medium	No	Excellent	No

Table 1: Class-labeled Training Tuples from the AllElectronics Customer Database

2 Entropy of the Target Variable

The entropy of the target variable (Buys Computer) is calculated as:

$$H(S) = -p_{yes}\log_2(p_{yes}) - p_{no}\log_2(p_{no})$$

Where:

$$p_{yes} = \frac{9}{14}, \quad p_{no} = \frac{5}{14}$$

$$H(S) = -\left(\frac{9}{14}\log_2\frac{9}{14}\right) - \left(\frac{5}{14}\log_2\frac{5}{14}\right) = 0.94$$

3 Entropy for Attribute Age

We now calculate the entropy for the attribute **Age** by splitting the dataset into Youth, Middle Aged, and Senior groups.

For Youth (5 instances):

$$H(Youth) = -\left(\frac{2}{5}\log_2\frac{2}{5}\right) - \left(\frac{3}{5}\log_2\frac{3}{5}\right) = 0.97$$

For Middle Aged (4 instances):

$$H(MiddleAged) = -\left(\frac{0}{4}\log_2\frac{0}{4}\right) - \left(\frac{4}{4}\log_2\frac{4}{4}\right) = 0.0$$

For Senior (5 instances):

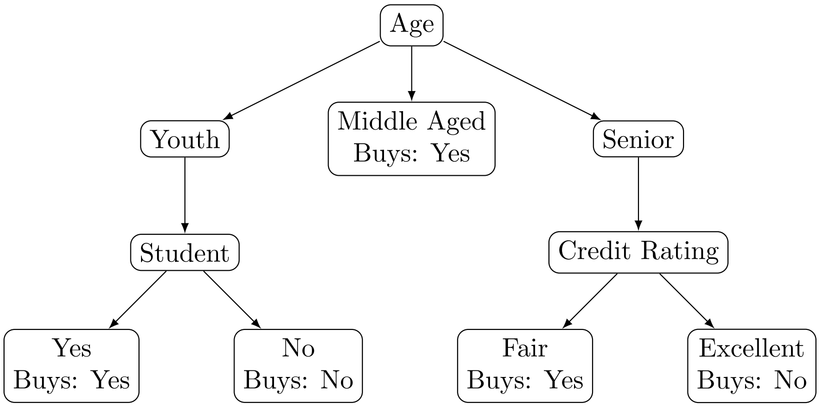
$$H(Senior) = -\left(\frac{3}{5}\log_2\frac{3}{5}\right) - \left(\frac{2}{5}\log_2\frac{2}{5}\right) = 0.97$$

The information gain for Age is calculated as:

$$\text{Gain}(S, \text{Age}) = 0.94 - \left(\frac{5}{14}\cdot 0.97 + \frac{4}{14}\cdot 0.0 + \frac{5}{14}\cdot 0.97\right) = 0.247$$

4 Decision Tree Diagram

Below is the diagram of the decision tree, where the root node is based on the attribute with the highest information gain.


Age
Youth
Middle Aged
Buys: Yes
Senior
Student
Credit Rating
Yes
Buys: Yes
No
Buys: No
Fair
Buys: Yes
Excellent
Buys: No