

CS483 - Fundamentals of Artificial Intelligence

Homework Assignment #4

Name: Shoumya Singh

Id: 19566

2. Given a dataset as follows, please buildup a decision tree with max information gain comparing the different condition checking features by hand calculation entropy and predict "Profit" in the new data. After that, write the Python program to verify your design

Age	Competition	Type	Profit
Old	Yes	Software	Down
Old	No	Software	Down
Old	No	Hardware	Down
Mid	Yes	Software	Down
Mid	Yes	Hardware	Down
Mid	No	Hardware	Up
Mid	No	Software	Up
New	Yes	Software	Up
New	No	Hardware	Up
New	No	Software	Up
Mid	No	Hardware	?

Based on these data, we can compute probability of Profit target.

- Since probability is equal to frequency relative, we have
 - Prob (Down) = 5 / 10
 - Prob (Up) = 5 / 10
- So now we calculate the Entropy using the below formula:

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$

$$= -(5/10) * \log_2(5/10) - (5/10) * \log_2(5/10)$$

$$= 1$$

1st Iteration: Find the root of a decision tree

The **Parent Data Table** has classes of 5 Down and 5 Up which produce entropy of 1.

- Information Gain for Age

Age	Profit
Old	Down
Old	Down
Old	Down
Mid	Down
Mid	Down
Mid	Up
Mid	Up
New	Up
New	Up
New	Up

Age	Profit
New	Up
New	Up
New	Up

Age	Profit
Old	Down
Old	Down
Old	Down

Age	Profit
Mid	Down
Mid	Down
Mid	Up
Mid	Up

$$\text{Entropy of New table} = -(3/3) * \log_2(3/3)$$

$$= 0$$

$$\text{Entropy of Old table} = -(3/3) * \log_2(3/3)$$

$$= 0$$

$$\text{Entropy of Mid table} = -(2/4) * \log_2(2/4) - (2/4) * \log_2(2/4) \\ = 1$$

The entropy of the Information Gain for Age

$$= 1 - (3/10 * 0 + 3/10 * 0 + 4/10 * 1) \\ = 0.6$$

- Information Gain for Competition

Competition	Profit
Yes	Down
No	Down
No	Down
Yes	Down
Yes	Down
No	Up
No	Up
Yes	Up
No	Up
No	Up

Competition	Profit
Yes	Down
Yes	Down
Yes	Down
Yes	Up

Competition	Profit
No	Down
No	Down
No	Up
No	Up
No	Up
No	Up

$$\text{Entropy of Yes table} = -(3/4) * \log_2(3/4) - (1/4) * \log_2(1/4) \\ = 0.8112$$

$$\text{Entropy of No table} = -(2/6) * \log_2(2/6) - (4/6) * \log_2(4/6) \\ = 0.9182$$

$$\text{The entropy of the Information Gain for Competition} \\ = 1 - (4/10 * 0.8112 + 6/10 * 0.9182) \\ = 0.1246$$

- Information Gain for Type

Type	Profit
Software	Down
Software	Down
Hardware	Down
Software	Down
Hardware	Down
Hardware	Up
Software	Up
Software	Up
Hardware	Up
Software	Up

Type	Profit
Software	Down
Software	Down
Software	Down
Software	Up
Software	Up
Software	Up

Type	Profit
Hardware	Down
Hardware	Down
Hardware	Up
Hardware	Up

$$\text{Entropy of Software table} = -(3/6) * \log_2(3/6) - (3/6) * \log_2(3/6) \\ = 1$$

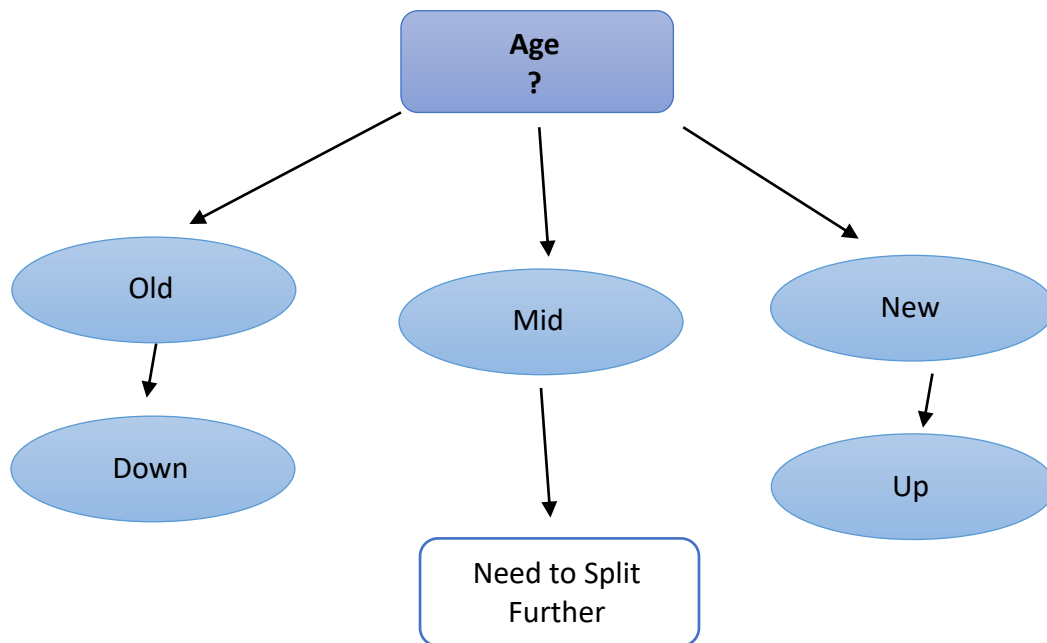
$$\text{Entropy of Hardware table} = -(2/4) * \log_2(2/4) - (2/4) * \log_2(2/4) \\ = 1$$

$$\text{The entropy of the Information Gain for Type} \\ = 1 - (6/10 * 1 + 4/10 * 1) \\ = 0$$

Table below summarizes the information gain for all four attributes.
Result of the First Iteration:

Gain	Age	Competition	Type
Entropy	0.6	0.1246	0

- Age is selected as the root because it has the highest information gain.



Now we have the maximum Information gain for feature Age, so we select it first.

Then,

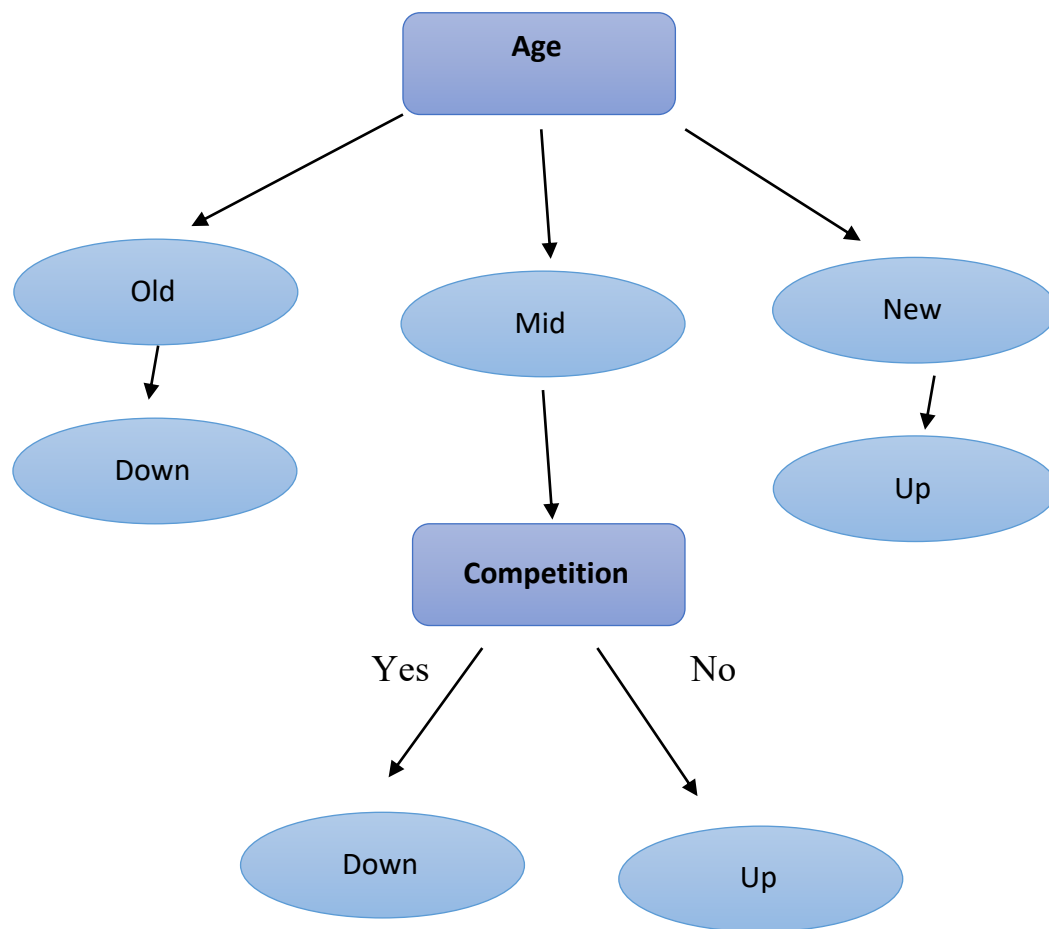
$$E(\text{Competition} = \text{Yes} \mid \text{Age} = \text{Mid}) = -1 \log 1 = 0 \text{ (Down)}$$

$$E(\text{Competition} = \text{No} \mid \text{Age} = \text{Mid}) = -1 \log 1 = 0 \text{ (Up)}$$

$$IG(\text{Competition} \mid \text{Age} = \text{Mid}) = 1 - 0 - 0 = 1$$

Now no need to find the other feature because this is the maximum possible IG for given Age = Mid So feature Competition is selected next.

Decision Tree:



So, for Prediction Age = Mid and Competition = No Then the Profit is Up.