BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI – HYDERABAD CAMPUS
BITS F464 : MACHINE LEARNING
FIRST SEMESTER 2023-2024

A worked-out example on Decision Trees

Here's a step-by-step solved numerical problem on the decision tree learning algorithm.

Let's assume the following dataset to classify whether a person will buy a computer based on two features: "Age" and "Income":

PERSON	AGE	INCOME	WILL BUY COMPUTER?
1	Young	Low	No
2	Young	Medium	Yes
3	Middle-aged	High	Yes
4	Middle-aged	Medium	Yes
5	Elderly	Medium	Yes
6	Elderly	High	No

We want to create a Decision Tree to classify whether a person will buy a computer or not based on their age and income.

Step 1: Calculate Entropy $H(D)$ of the Dataset:

Calculate the entropy of the dataset based on the target variable "Will Buy Computer".

$$H(D) = -\left[\, p_{Yes} \log_2(p_{Yes}) + p_{No} \log_2(p_{No}) \,\right]$$

where p_{Yes} is the proportion of "Yes" instances and p_{No} is the proportion of "No" instances.

For this dataset:

$$p_{Yes} = \frac{4}{6} = 0.667$$
$$p_{No} = \frac{2}{6} = 0.333$$

$$H(D) = -0.667 \log_2(0.667) - 0.333 \log_2(0.333) = 0.918$$

Step 2: Calculate Information Gain IG for Each Feature:

Calculate the information gain for each feature by calculating the weighted average of entropy after splitting the dataset based on that feature.

$$IG(D, X) = H(D) - \sum_{v \,\epsilon\, Values(X)} \frac{|D_v|}{|D|} H(D_v)$$

where X is the feature being considered, $Values(X)$ are the possible values of the feature, $|D_v|$ is the number of instances with value v in feature X, and $H(D_v)$ is the entropy of the subset D_v.

For “Age”:

- Values(“Age”): {Young, Middle-aged, Elderly}
- Entropy(“Age” = Young): $H(D_{Young})=1$
- Entropy(“Age” = Middle-aged): $H(D_{Middle\text{-}aged})=0$
- Entropy(“Age” = Elderly): $H(D_{Elderly})=1$

$$IG(D, Age) = 0.918 - \frac{2}{6}\cdot 1 - \frac{2}{6}\cdot 0 - \frac{2}{6}\cdot 1 = 0.251$$

For “Income”:

- Values(“Income”): {Low, Medium, High}
- Entropy(“Income” = Low): $H(D_{Low})=0$
- Entropy(“Income” = Medium): $H(D_{Medium})=0$
- Entropy(“Income” = High): $H(D_{High})=1$

$$IG(D, Income) = 0.918 - \frac{1}{6}\cdot 0 - \frac{3}{6}\cdot 0 - \frac{2}{6}\cdot 1 = 0.585$$

Step 3: Choose the Feature with the Highest Information Gain:

Choose the feature that has the highest information gain as the root of the decision tree. In this case, “Income” has the highest information gain ($IG=0.585$).

Step 4: Split the Dataset and Recurse:

Split the dataset based on the chosen feature and repeat the decision tree learning process for each subset.

For the subset with “Income” = Low, all instances have the same target value (“No”). Thus, the decision tree node is a leaf node labeled “No”.

For the subset with “Income” = Medium, all instances have the same target value (“Yes”). Thus, the decision tree node is a leaf node labeled “Yes”.

For the subset with “Income” = High, there are mixed target values (“Yes” and “No”). We will continue the decision tree learning process for this subset.

Step 5: Calculate Information Gain for Features in Subsets:

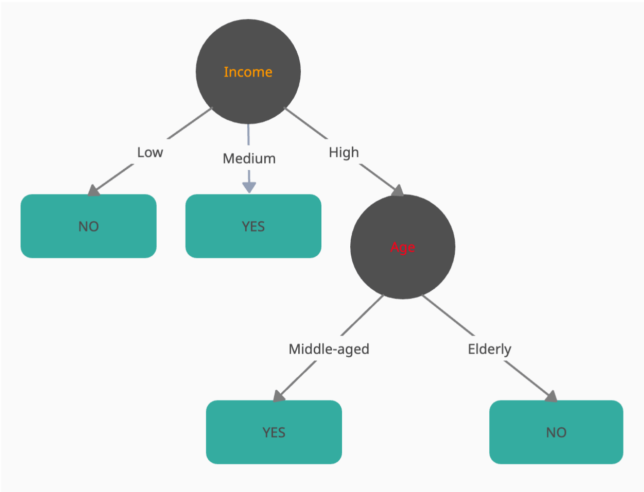
For the subset with “Income” = High, calculate the information gain for each feature (“Age”).

For “Age”:

- Values(“Age”): {Middle-aged, Elderly}
- Entropy(“Age” = Middle-aged): $H(D_{Middle\text{-}aged})=0$
- Entropy(“Age” = Elderly): $H(D_{Elderly})=0$

Since there is only one feature left to consider, we choose “Age” as the next node of the decision tree, with one leaf node “Age = Middle-aged” labeled “Yes” and another leaf node “Age = Elderly” labeled “No”.

The final decision tree:



This concludes the step-by-step example of the Decision Tree learning algorithm. The algorithm selects the features that maximize information gain at each step to build a decision tree for classification.