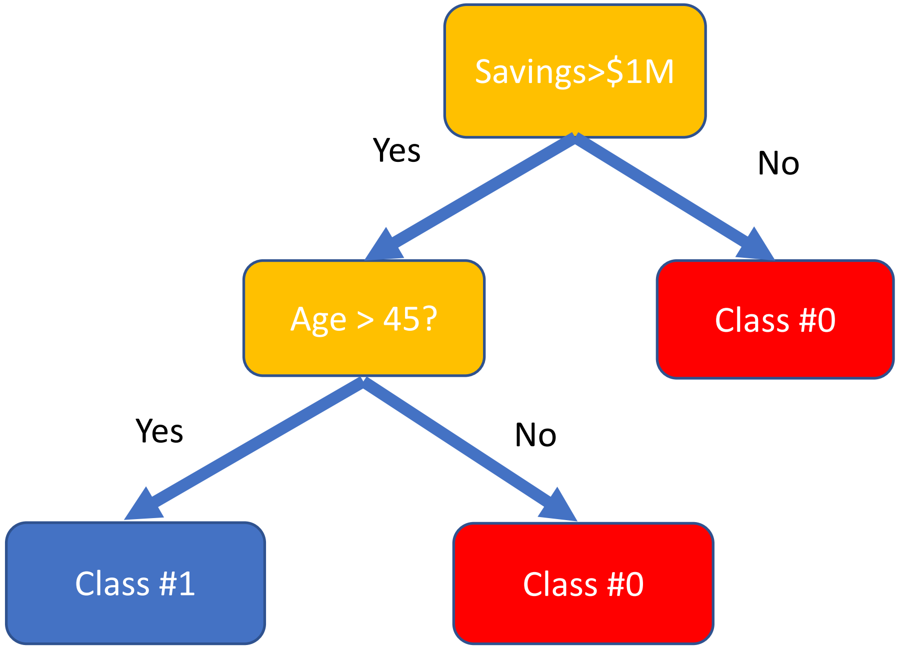
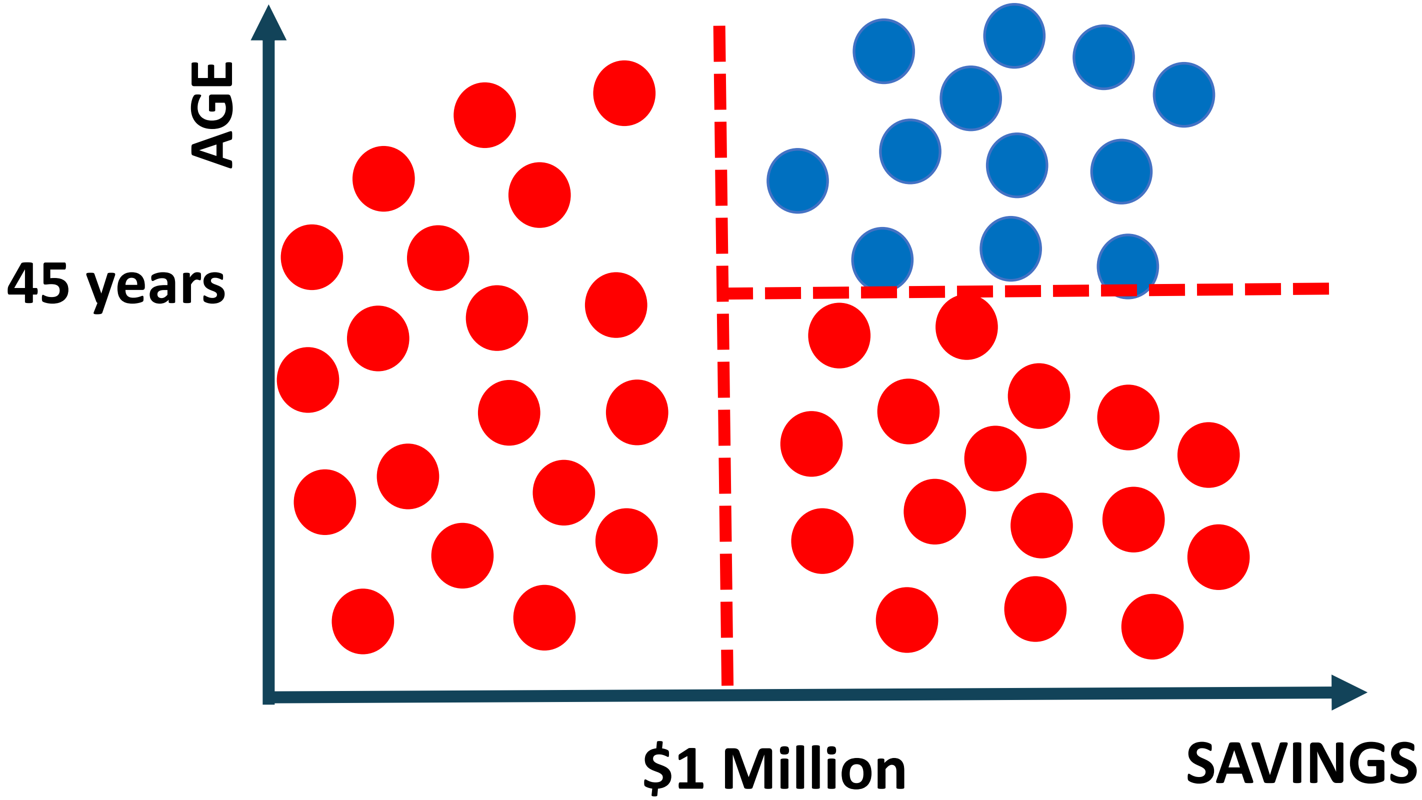
Machine Learning

06048203

DECISION TREES: INTUITION

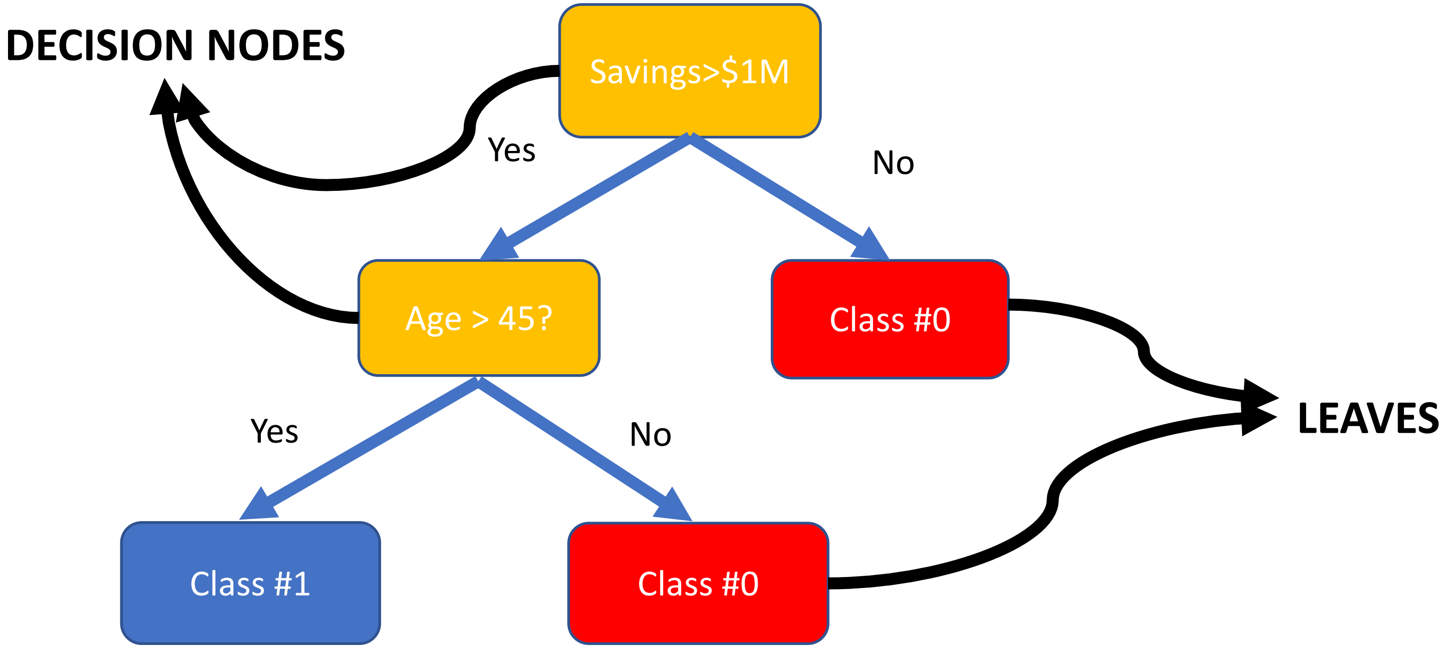
- Decision Trees are supervised Machine Learning technique where the data is split according to a certain condition/parameter.
- Let’s assume we want to classify whether a customer could retire or not based on their savings and age.



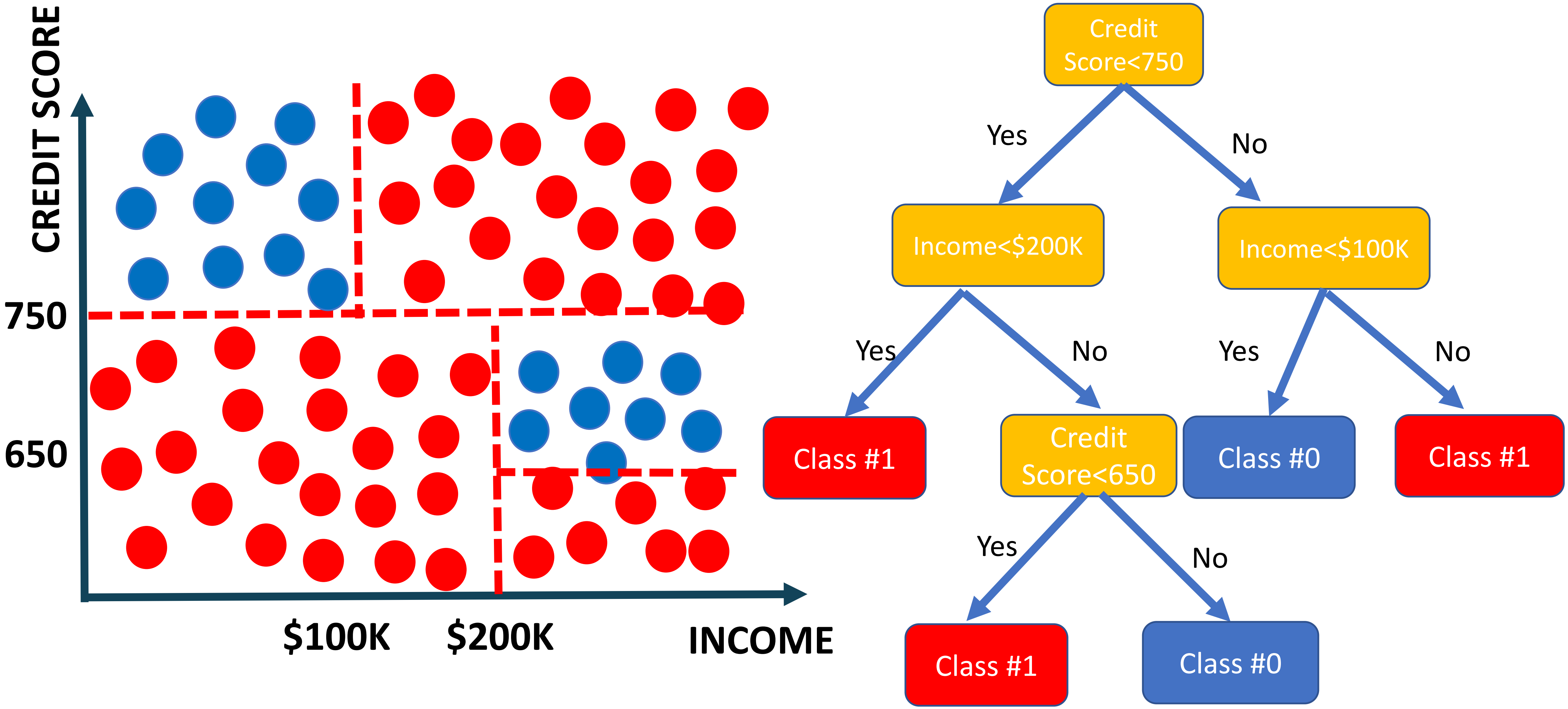


DECISION TREES: DEFINITIONS

- The tree consists of **decision nodes** and **leaves**.
- Leaves are the decisions or the final outcomes.
- Decision nodes are where the data is split based on a certain attribute.
- Objective is to minimize the entropy which provides the optimum split



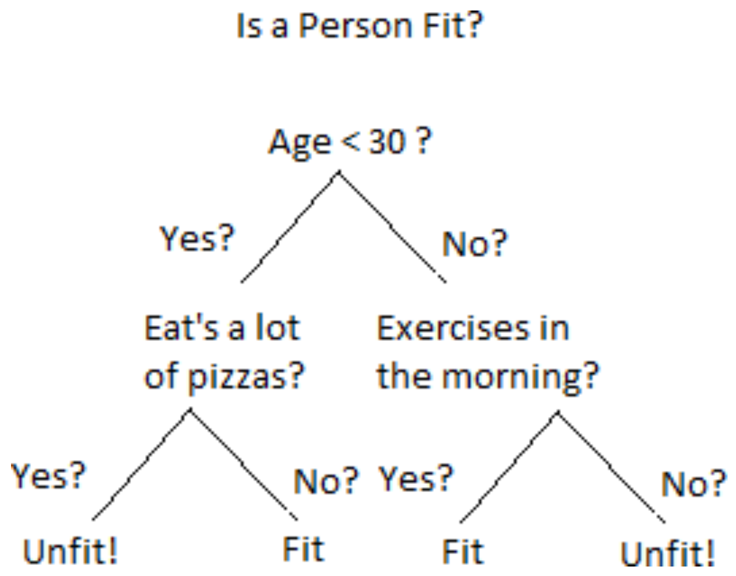
DECISION TREES: CUSTOMER SEGMENTATION



Decision Trees - Basics

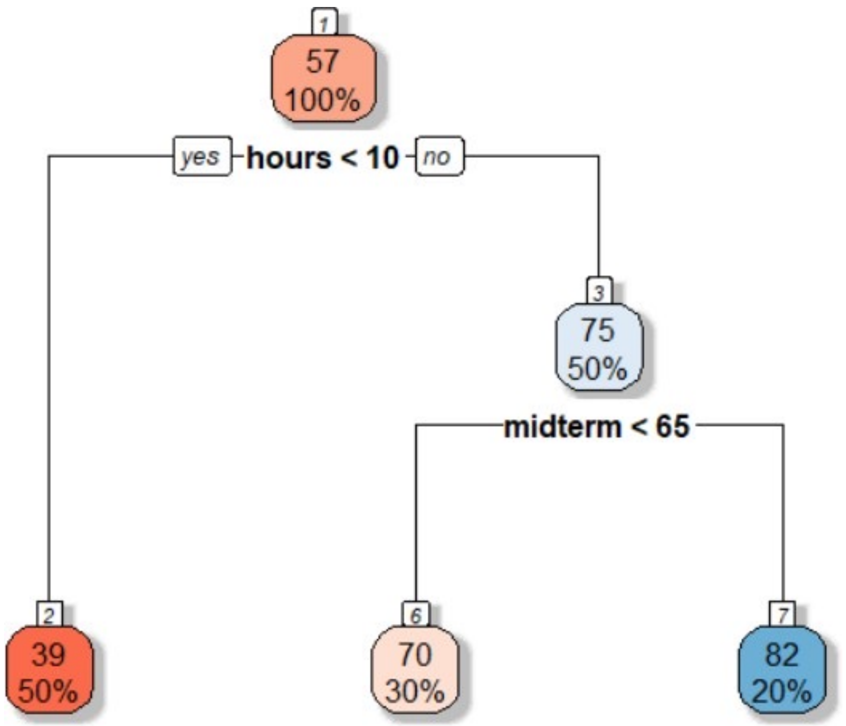
Definition

A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.



Example

	score	hours	midterm
1	35	6	42
2	38	5	65
3	40	7	35
4	45	6	75
5	35	8	60
6	65	11	50
7	70	12	45
8	75	18	40
9	80	14	80
10	85	12	82



Decision Trees - Basics

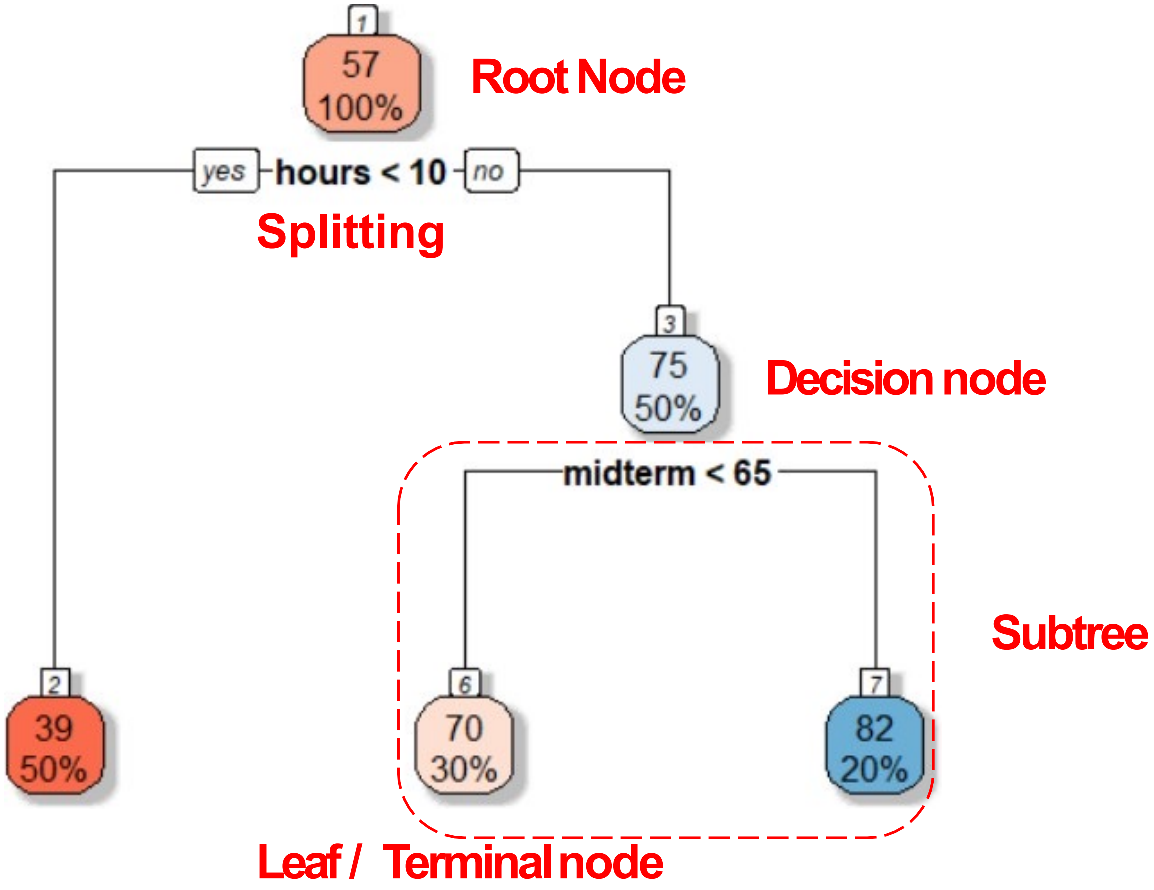
Types

1. Regression Tree
   For continuous quantitative target variable.
   Eg. Predicting rainfall, predicting revenue, predicting marks etc.

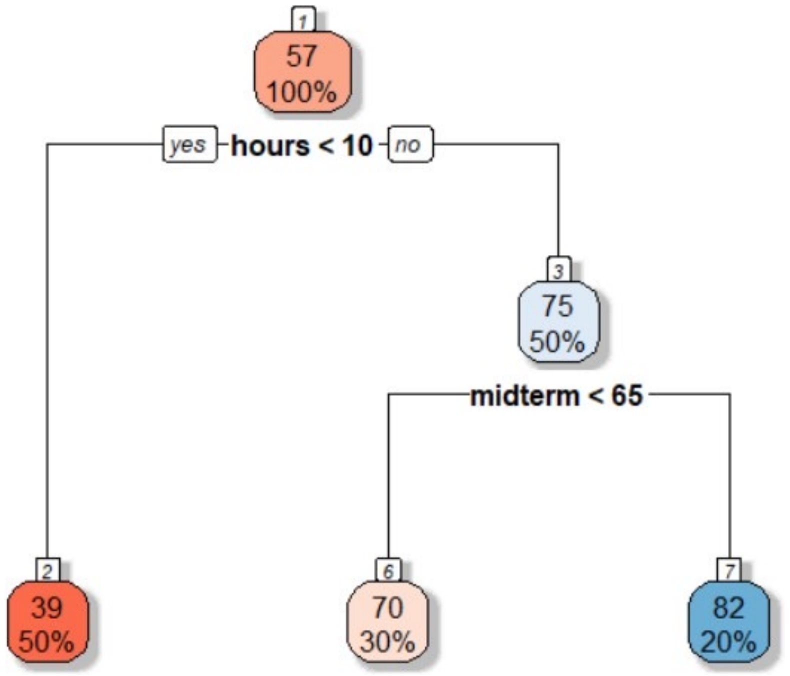
2. Classification Tree
   For discrete categorical target variables
   Eg. Predicting High or Low, Win or Loss, Healthy or Unhealthy etc

Decision Trees - Basics

Terminologies



Steps



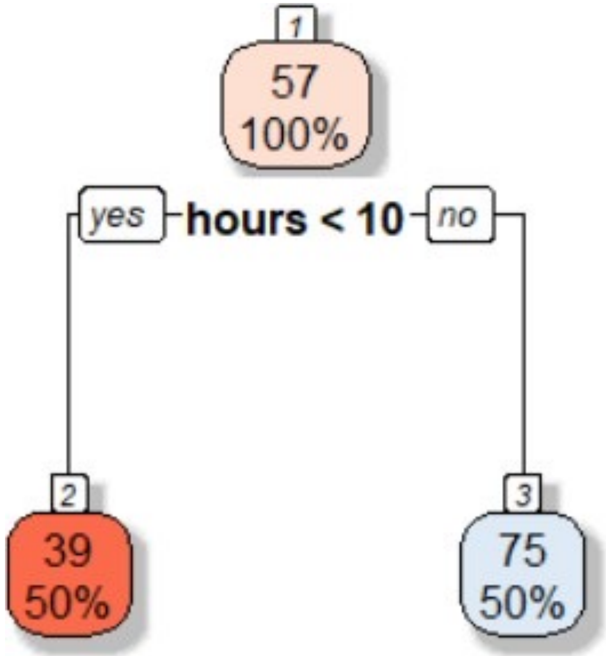
1. We divide the predictor space—that is, the set of possible values for X1,X2, . . .,Xp—into J distinct and non-overlapping regions, R1,R2, . . . , RJ .

2. For every observation that falls into the region Rj, we make the same prediction, which is simply the mean of the response values for the training observations in Rj .

Goal is to minimize RSS

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

Building tree



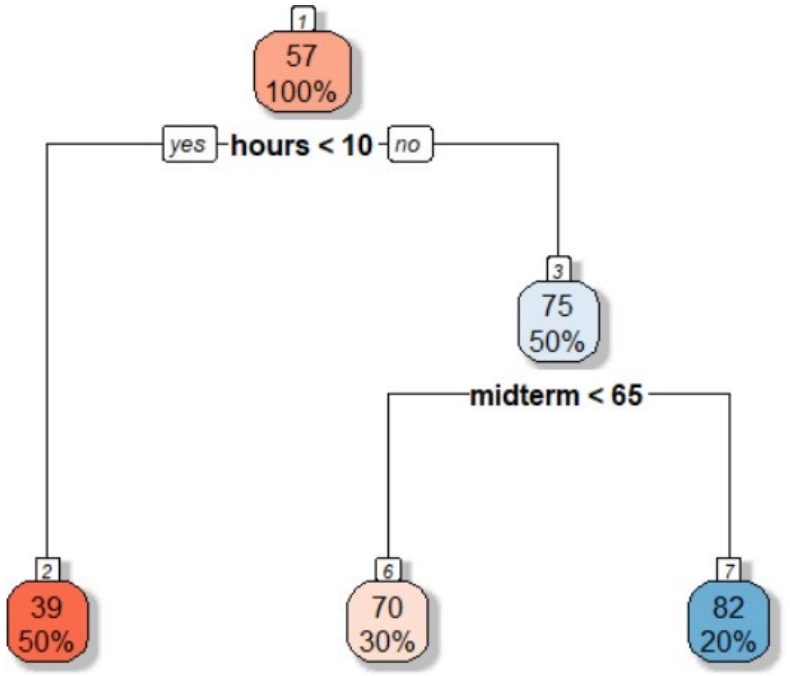
	score	hours	midterm
1	35	6	42
2	38	5	65
3	40	7	35
4	45	6	75
5	35	8	60
6	65	11	50
7	70	12	45
8	75	18	40
9	80	14	80
10	85	12	82

Mean score 39

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2$$

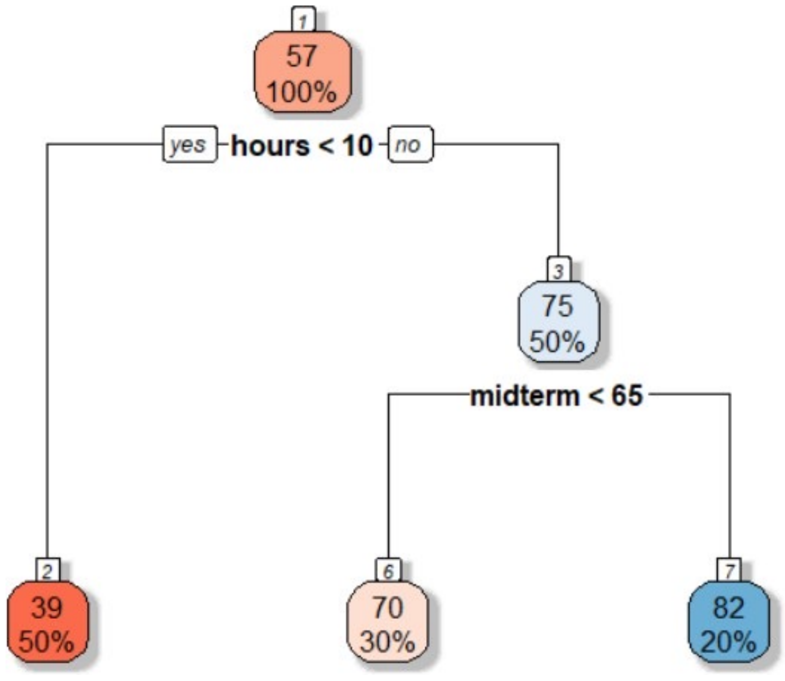
Mean score 75

Approach



- Top-down, greedy approach that is known as recursive binary splitting.
- Top-down because it begins at the top of the tree and then successively splits the predictor space
- Each split is indicated via two new branches further down on the tree.
- It is greedy because at each step of the tree-building process, the best split is made at that particular step, rather than looking ahead and picking a split that will lead to a better tree in some future step.

Steps



1. Considers all predictors and all possible cut point values
2. Calculates RSS for each possibility
3. Selects the one with least RSS
4. Continues till stopping criteria is reached

Decision Trees

Types

1. Regression Tree
   For continuous quantitative target variable.
   Eg. Predicting rainfall, predicting revenue, predicting marks etc.

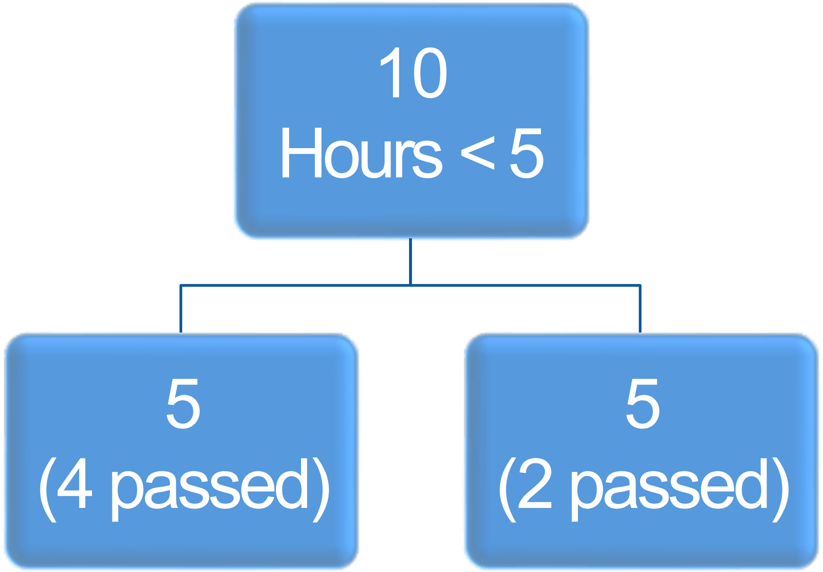
2. Classification Tree
   For discrete categorical target variables
   Eg. Predicting High or Low, Win or Loss, Healthy or Unhealthy etc

Classification Trees

Prediction method

Regression
Mean of response variable became prediction for that class

Classification
We use mode (most frequent category in that region will be the prediction)



Classification Trees

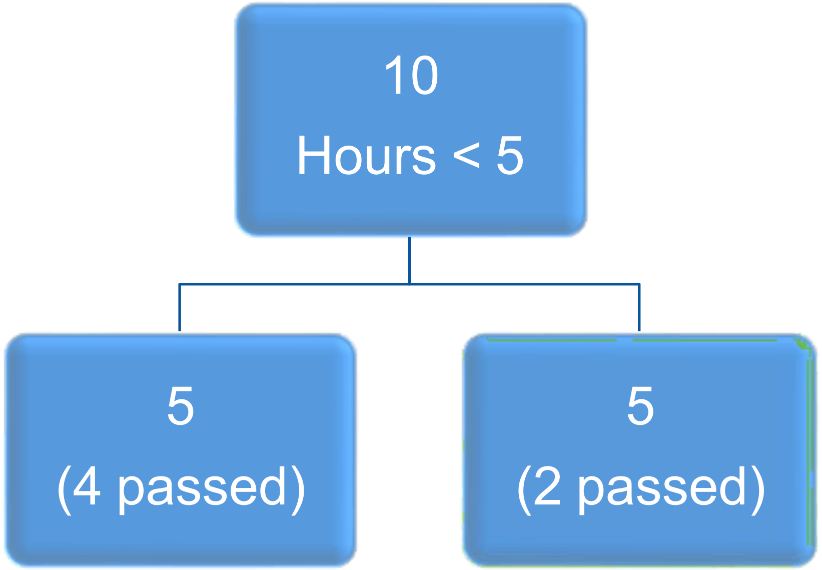
Methods

Both Regression and classification use recursive binary splitting

In Regression RSS is used to decide the split

In Classification we can use

1. Classification error rate
2. Gini Index
3. Cross Entropy



Classification Trees

Methods

In Classification we can use

1. Classification error rate
2. **Gini Index**
3. **Cross Entropy**

Gini index and cross entropy signifies node purity

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

$$D = -\sum_{k=1}^{K} \hat{p}_{mk} \log \hat{p}_{mk}$$