



Start-Tech Academy

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

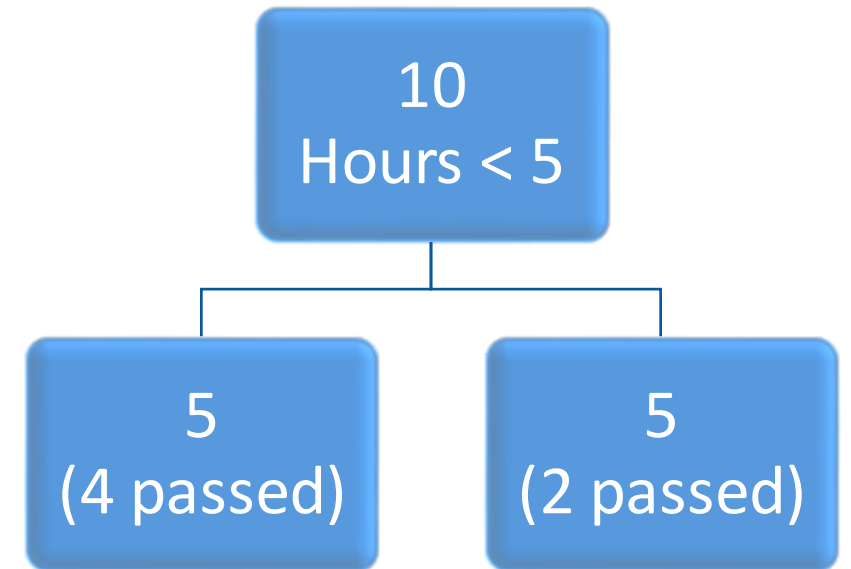
Prediction method

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

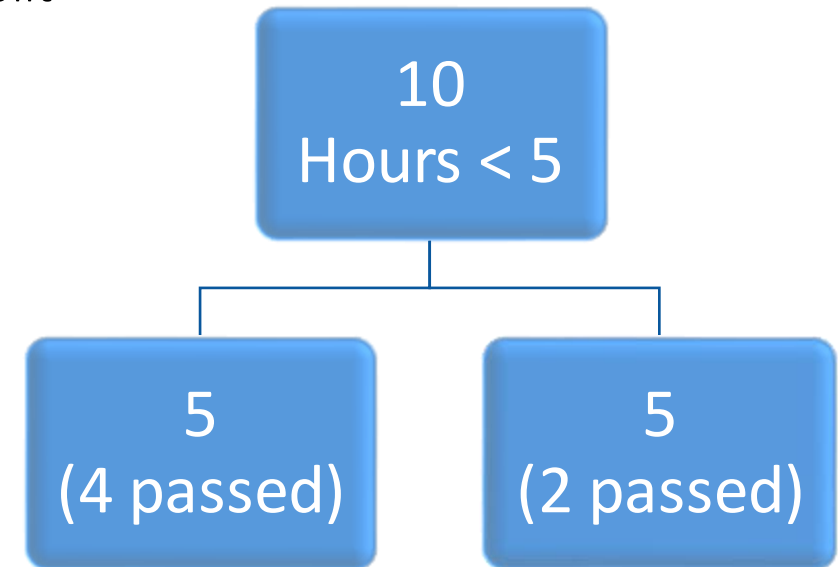
Methods

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

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

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

