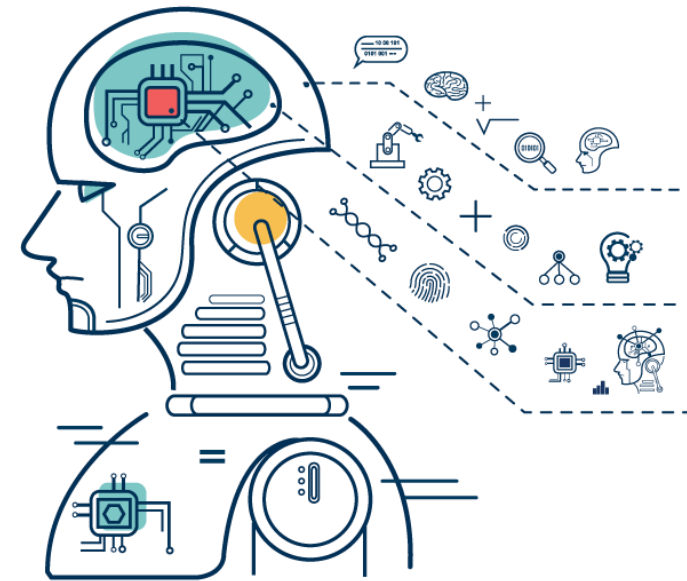


Siddhardhan

Model Selection in Machine Learning

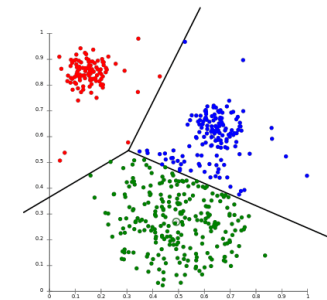
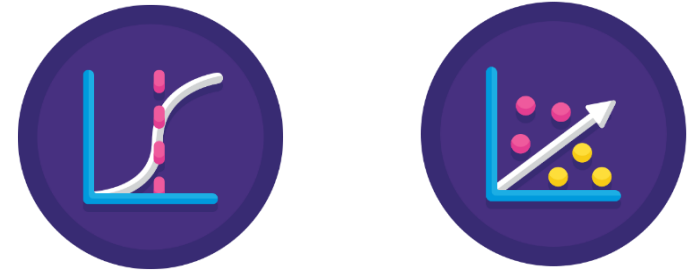


Model Selection

Model Selection in Machine Learning is the process of choosing the best suited model for a particular problem. Selecting a model depends on various factors such as the dataset, task, nature of the model, etc.

Two factors to be considered:

1. Logical Reason to select a model
2. Comparing the performance of the models



Model Selection

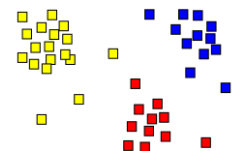
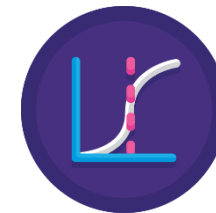
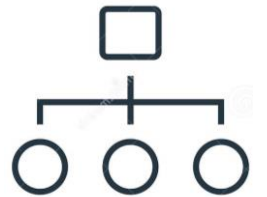
Models can be selected based on :

1. Type of Data available:

- a. Images & Videos – CNN
- b. Text data or Speech data – RNN
- c. Numerical data – SVM, Logistic Regression, Decision trees, etc.

2. Based on the task we need to carry out:

- a. Classification tasks – SVM, Logistic Regression, Decision trees, etc.
- b. Regression tasks – Linear regression, Random Forest, Polynomial regression, etc.
- c. Clustering tasks – K-Means Clustering, Hierarchical Clustering



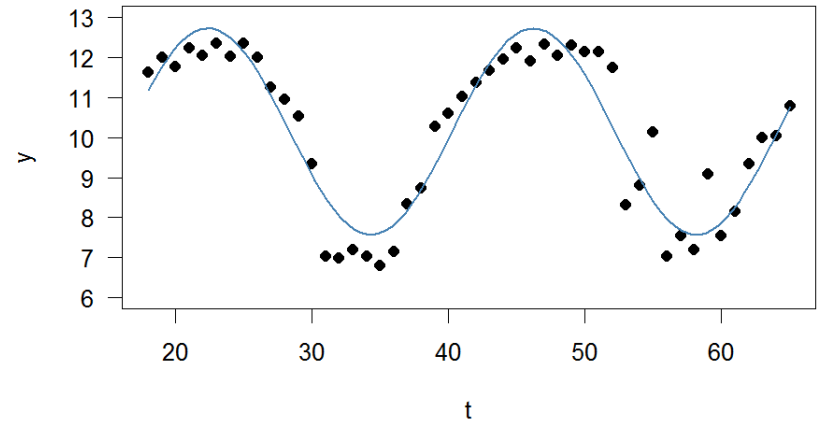
Linear Regression

Advantages:

1. Very simple to implement
2. Performs well on data with linear relationship

Disadvantages:

1. Not suitable for data having non-linear relationship
2. Underfitting issue
3. Sensitive to Outliers



Logistic Regression

Advantages:

1. Easy to implement
2. Performs well on data with linear relationship
3. Less prone to over-fitting for low dimensional dataset

Disadvantages:

1. High dimensional dataset causes over-fitting
2. Difficult to capture complex relationships in a dataset
3. Sensitive to Outliers
4. Needs a larger dataset



Decision Tree

Advantages:

1. Can be used for both Classification & Regression
2. Easy to interpret
3. No need for normalization or scaling
4. Not sensitive to outliers

Disadvantages:

1. Overfitting issue
2. Small changes in the data alter the tree structure causing instability
3. Training time is relatively higher

