

Linear Regression:

linear regression is a supervised and statistical model used to predict the relationship between independent and dependent variables and It is used when we want to predict the value of a dependent variable based on the value of independent variable.

This kind of model examine two factors:

- 1- Which variable in particular are significant predictors of the outcome variables.
- 2- How significant is the regression line to make predictions with possible accuracy (if it's inaccurate we can't use it so it's very important we found out the most accurate line we can get).

There are some applications of linear regression that contains:

- 1- Economic Growth: Used to determine the economic Growth of a country or a state in to coming quarter, can also be used to predict the GDP of a country.
- 2- Product Price: Can be used to predict what would be the price of a product in the future.
- 3- Housing Sales: To estimate the number of houses a builder would sell and at what price in the coming months.
- 4- Score Prediction: To predict the number of runs a player would score in the coming matches based on previous performance.

Decision Tree:

Decision tree, which is a supervised learning method, is a tree shaped diagram used to determine a course of actions. Each branch of the tree represents a possible decision, occurrence or reaction.

Problems that decision tree can solve:

- 1- Classification: A classification tree will determine a set of logical if-then conditions to classify problems. For example, discriminating between three type of flowers based on certain features.
- 2- Regression: Regression tree is used when the largest variable is numerical or continuous in nature. We fit a regression model to the target variable using each of the independent variables. Each split is made based on the sum of squared error.

Advantages of decision tree contains:

- 1- Simple to understand and visualize.
- 2- Little effort required for data preparation.
- 3- Can handle both numerical and categorical data.
- 4- Nonlinear parameters don't affect its performance.

Disadvantages of decision tree includes:

- 1- Overfitting: overfitting occurs when the algorithm capture noise in the data.
- 2- High variance: The model can get unstable due to small variation in data.

- 3- Low biased tree: A highly complicated decision tree tends to have a low bias which make it difficult for the model to work with new data.

KNN Algorithm:

K nearest neighbour is the simplest supervised machine learning algorithm mostly used for classification. It classifies a data point based on how its neighbours are classified. KNN stores all available cases and classifies new case based on a similarity measure. K in KNN is a parameter that refers to the number of nearest neighbour to include in the majority voting process. A data point is classified by the majority votes from its K nearest neighbours.

KNN is based on feature similarity. Choosing the right value of K is a process called parameter tuning, and is important for better accuracy. To choose a value of K:

- 1- Square(n), where n is the total number of data points.
- 2- Odd value of k is selected to avoid confusion between to classes of data.

We can use KNN when:

- 1- Data is simple.
- 2- Data is noise free.
- 3- Dataset is small (because KNN is a lazy learner i.e. doesn't learn a discriminative function from the training set.

How does the algorithm works?

Consider a dataset having two variables: height(cm) and weight(kg) and each point is classified as normal or underweight. On the basis of the given data we have to classify the set as normal or underweight using KNN. To find nearest neighbour, we will calculate Euclidean distance. Hence, we have calculated the Euclidean distance of unknown data point from all the points, and select the K entries in our database which are closest to the new sample. Then we find the most common classification of this entries, this is the classification we give to the new sample.