**Machine Learning Algorithms**

# Linear Regression in Machine Learning

Linear regression is one of the most accessible and popular Machine Learning algorithms. It is a statistical method that is used for predictive analysis. For example, linear regression makes predictions for continuous/actual or numeric variables such as **sales, salary, age, product price,** etc.

Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (y) variables, hence called linear regression. Since linear regression shows the linear relationship, it finds how the value of the dependent variable is changing according to the value of the independent variable.

## **Types of Linear Regression**

Linear regression can be further divided into two types of algorithm:

**Simple Linear Regression:**

If a single independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Simple Linear Regression.

Multiple Linear regression:

If more than one independent variable is used to predict the value of a numerical dependent variable, then such a Linear Regression algorithm is called Multiple Linear Regression.

# Advantages of Linear Regression

### **Simple implementation**

Linear Regression is a straightforward algorithm that can quickly be implemented to give satisfactory results. Furthermore, these models can be trained rapidly and efficiently even on systems with relatively low computational power compared to other complex algorithms. Linear regression has a considerably lower time complexity when compared to some of the different machine learning algorithms. The mathematical equations of Linear regression are also reasonably easy to understand and interpret. Hence Linear regression is straightforward to master.

### **Performance on linearly separable datasets**

Linear regression fits linearly separable datasets almost perfectly and is often used to find the nature of the relationship between variables.

### **Overfitting can be reduced by regularization**

Overfitting is a situation that arises when a machine learning model fits a dataset very closely and hence captures the noisy data as well; this negatively impacts the performance of the model and reduces its accuracy on the test set.  
Regularization is a technique that can be easily implemented and can effectively reduce the complexity of a function to reduce the risk of overfitting.

# Disadvantages of Linear RegressionProne to underfitting

**Underfitting**: A situation that arises when a machine learning model fails to capture the data correctly; typically, the hypothesis function cannot fit the data well.

Since linear regression assumes a linear relationship between the input and output variables, it fails to fit complex datasets properly; in most real-life scenarios, the relationship between the dataset variables isn't linear, and hence a straight line doesn't fit the data correctly. A more complex function can capture the data more effectively in such situations. Because of this, most linear regression models have low accuracy.

### **Sensitive to outliers**

Outliers of a data set are anomalies or extreme values that deviate from the other data points of the distribution. Data outliers can damage the performance  
of a machine learning model drastically and can often lead to models with low  
Accuracy.

**K-Nearest Neighbor(KNN) Algorithm for Machine Learning**

* K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on the Supervised Learning technique.
* The K-NN algorithm assumes the similarity between the new and available cases and puts the new data into the most similar category to the available types.
* The K-NN algorithm stores all the available data and classifies a new data point based on the similarity; this means when new data appears, it can be easily classified into a good suite category by using the K- NN algorithm.
* The K-NN algorithm can be used for Regression and Classification, but mainly for Classification problems.
* K-NN is a non-parametric algorithm, which means it does not make any assumptions on underlying data.
* It is also called a lazy learner algorithm because it does not immediately learn from the training set; instead, it stores the dataset. Then, at the time of Classification, it acts on the dataset.
* KNN algorithm at the training phase just stores the dataset, and when it gets new data, it classifies that data into a much similar category to the latest data.
* **Example:** Suppose we have an image of a creature that looks similar to a cat and dog, but we want to know whether it is a cat or a dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. So our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features, it will put it in either cat or dog category.

## 

## **How does K-NN work?**

The K-NN working can be explained based on the below algorithm:

* **Step-1:** Select the number K of the neighbors
* **Step-2:** Calculate the Euclidean distance of **K number of neighbors**
* **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.
* **Step-4:** Among these k neighbors, count the number of the data points in each category.
* **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
* **Step-6:** Our model is ready.

## **Advantages of KNN Algorithm:**

* It is simple to implement.
* It is robust to the noisy training data
* It can be more effective if the training data is extensive.

## **Disadvantages of KNN Algorithm:**

* Always needs to determine the value of K, which may be complex sometimes.
* The computation cost is high because of calculating the distance between the data points for all the
* training samples.