Linear Models for Classification

Classification:

- Classification is a process of finding a function which helps in dividing the dataset into classes based on different parameters.
- In Classification, a computer program is trained on the training dataset and based on that training, it categorizes the data into different classes.

Example:

- The best example to understand the Classification problem is Email Spam Detection.
- The model is trained on the basis of millions of emails on different parameters, and whenever it receives a new email, it identifies whether the email is spam or not.
- If the email is spam, then it is moved to the Spam folder.

Types of ML Classification Algorithms:

Classification Algorithms can be further divided into the following types:

- Logistic Regression
- K-Nearest Neighbours
- Support Vector Machines
- Kernel SVM
- Naïve Bayes
- Decision Tree Classification
- Random Forest Classification

Regression:

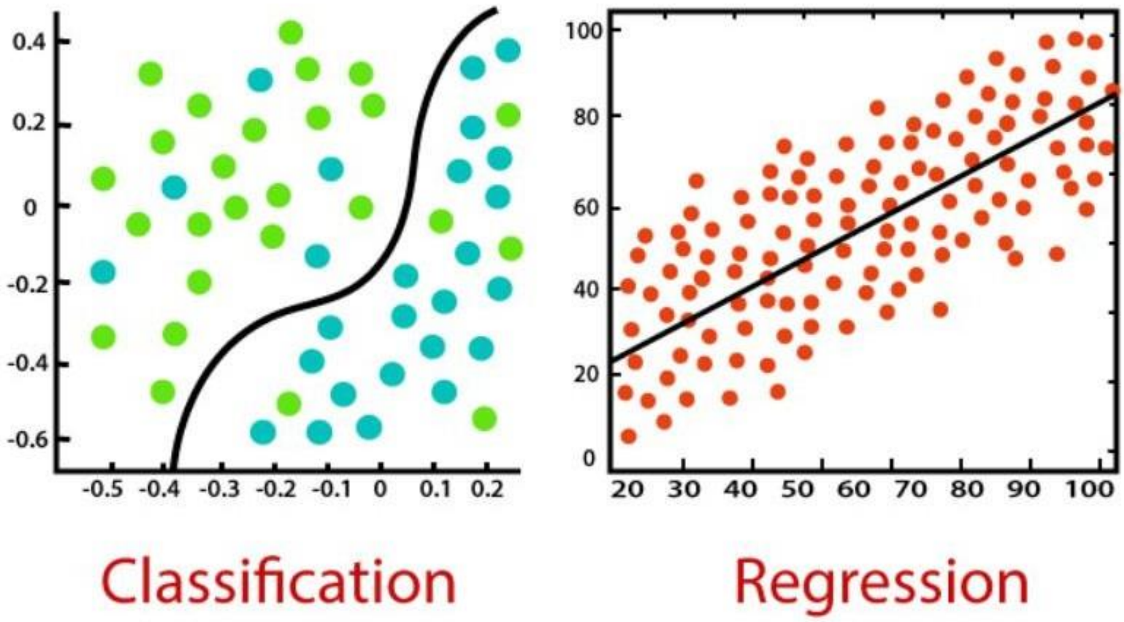
- Regression is a process of finding the correlations between dependent and independent variables.
- It helps in predicting the continuous variables such as prediction of Market Trends, prediction of House prices, etc.

- The task of the Regression algorithm is to find the mapping function to map the input variable(x) to the continuous output variable(y).

Example: Suppose we want to do weather forecasting, so for this, we will use the Regression algorithm. In weather prediction, the model is trained on the past data, and once the training is completed, it can easily predict the weather for future days.

Types of Regression Algorithm:

- Simple Linear Regression
- Multiple Linear Regression
- Polynomial Regression
- Support Vector Regression
- Decision Tree Regression
- Random Forest Regression



Linear Discriminant Analysis

1. Linear Discriminant Analysis (LDA) is a supervised learning algorithm used for classification tasks in machine learning. It is a technique used to find a linear combination of features that best separates the classes in a dataset.
2. LDA works by projecting the data onto a lower-dimensional space that maximizes the separation between the classes. It does this by finding a set of linear discriminants that maximize the ratio of between-class variance to within-class variance. In other words, it finds the directions in the feature space that best separate the different classes of data.
3. LDA assumes that the data has a Gaussian distribution and that the covariance matrices of the different classes are equal. It also assumes that the data is linearly separable, meaning that a linear decision boundary can accurately classify the different classes.

LDA has several advantages, including:

- It is a simple and computationally efficient algorithm.
- It can work well even when the number of features is much larger than the number of training samples.
- It can handle multicollinearity (correlation between features) in the data.

However, LDA also has some limitations, including:

- It assumes that the data has a Gaussian distribution, which may not always be the case.
- It assumes that the covariance matrices of the different classes are equal, which may not be true in some datasets.
- It assumes that the data is linearly separable, which may not be the case for some datasets.
- It may not perform well in high-dimensional feature spaces.

- Linear Discriminant Analysis or Normal Discriminant Analysis or Discriminant Function Analysis is a dimensionality reduction technique that is commonly used for supervised classification problems.
- It is used for modelling differences in groups i.e. separating two or more classes.
- It is used to project the features in higher dimension space into a lower dimension space.

  For example, we have two classes and we need to separate them efficiently.
- Classes can have multiple features. Using only a single feature to classify them may result in some overlapping as shown in the below figure.
- So, we will keep on increasing the number of features for proper classification.

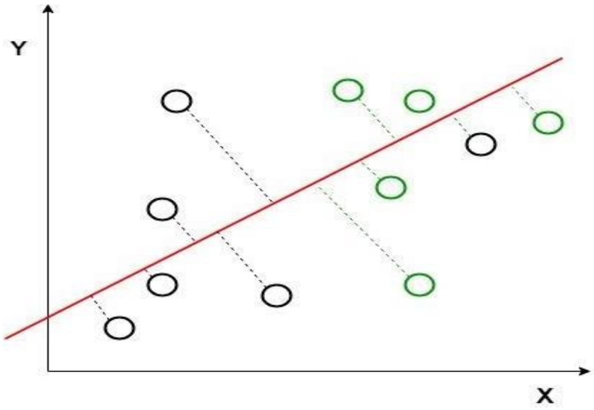
Example:

Suppose we have two sets of data points belonging to two different classes that we want to classify. As shown in the given 2D graph, when the data points are plotted on the 2D plane, there's no straight line that can separate the two classes of the data points completely. Hence, in this case, LDA (Linear Discriminant Analysis) is used which reduces the 2D graph into a 1D graph in order to maximize the separability between the two classes.

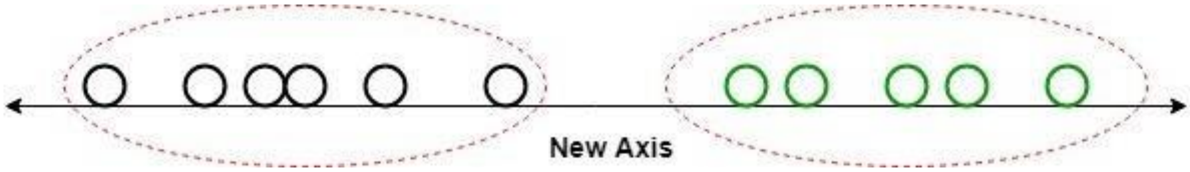
Here, Linear Discriminant Analysis uses both the axes (X and Y) to create a new axis and projects data onto a new axis in a way to maximize the separation of the two categories and hence, reducing the 2D graph into a 1D graph.

Two criteria are used by LDA to create a new axis:

1. Maximize the distance between means of the two classes.
2. Minimize the variation within each class.



In the above graph, it can be seen that a new axis (in red) is generated and plotted in the 2D graph such that it maximizes the distance between the means of the two classes and minimizes the variation within each class. In simple terms, this newly generated axis increases the separation between the data points of the two classes. After generating this new axis using the above-mentioned criteria, all the data points of the classes are plotted on this new axis and are shown in the figure given below.



But Linear Discriminant Analysis fails when the mean of the distributions are shared, as it becomes impossible for LDA to find a new axis that makes both the classes linearly separable. In such cases, we use non-linear discriminant analysis.