



Unit-2 DA

Data Analytics (Dr. A.P.J. Abdul Kalam Technical University)



Scan to open on Studocu

UNIT-2

Regression modelling

A regression model provides a function that describes the relationship between one or more independent variables and a response, dependent, or target variable.

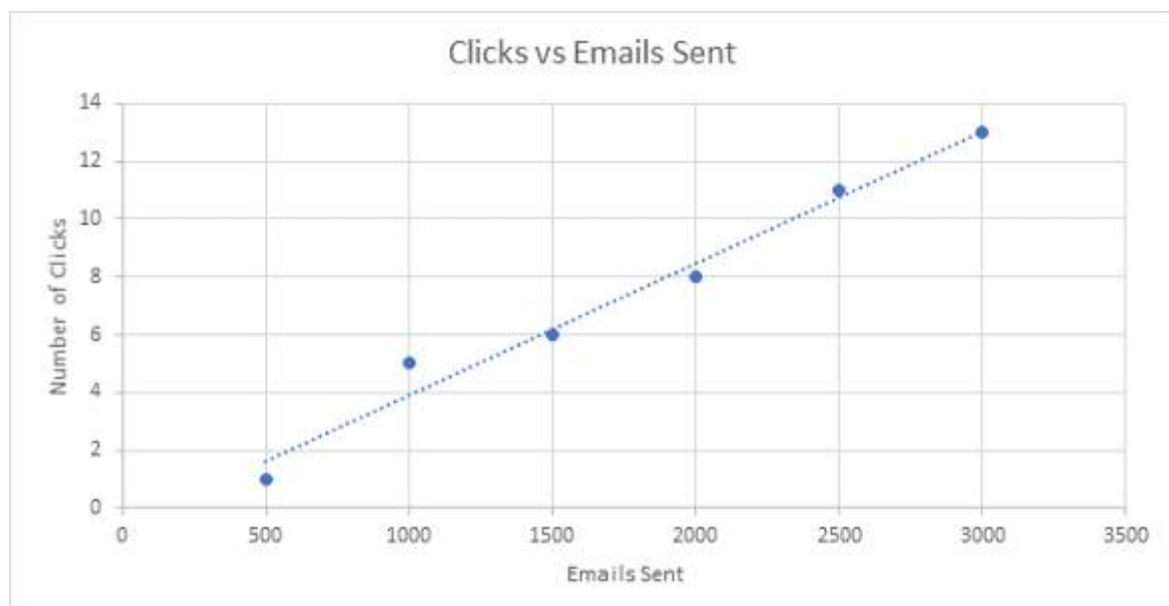
For example, the relationship between height and weight may be described by a linear regression model. A regression analysis is the basis for many types of prediction and for determining the effects on target variables. When you hear about studies on the news that talk about fuel efficiency, or the cause of pollution, or the effects of screen time on learning, there is often a regression model being used to support their claims.

Types of Regression

1. Linear

A linear regression is a model where the relationship between inputs and outputs is a straight line. This is the easiest to conceptualize and even observe in the real world. Even when a relationship isn't very linear, our brains try to see the pattern and attach a rudimentary linear model to that relationship.

One example may be around the number of responses to a marketing campaign. If we send 1,000 emails, we may get five responses. If this relationship can be modelled using a linear regression, we would expect to get ten responses when we send 2,000 emails. Your chart may vary, but the general idea is that we associate a predictor and a target, and we assume a relationship between the two.



Using a linear regression model, we want to estimate the correlation between the number of emails sent and response rates. In other words, if the linear model fits our observations well enough, then we can estimate that the more emails we send, the more responses we will get.

When making a claim like this, whether it is related to exercise, happiness, health, or any number of claims, there is usually a regression model behind the scenes to support the claim.

In addition, the model fit can be described using a mean squared error. This basically gives us a number to show exactly how well the linear model fits.

More serious examples of a linear regression would include predicting a patient's length of stay at a hospital, relationship between income and crime, education and birth rate, or sales and temperature.

2. Multiple

Multiple regression indicates that there are more than one input variables that may affect the outcome, or target variable. For our email campaign example, you may include an additional variable with the number of emails sent in the last month.

By looking at both input variables, a clearer picture starts to emerge about what drives users to respond to a campaign and how to optimize email timing and frequency. While conceptualizing the model becomes more complex with more inputs, the relationship may continue to be linear.

For these models, it is important to understand exactly what effect each input has and how they combine to produce the final target variable results.

3. Non-Linear

Nonlinear regression is a form of regression analysis in which data is fit to a model and then expressed as a mathematical function. Simple linear [regression](#) relates two variables (X and Y) with a straight line ($y = mx + b$), while nonlinear regression relates the two variables in a nonlinear (curved) relationship.

The goal of the model is to make the [sum of the squares](#) as small as possible. The sum of squares is a measure that tracks how far the Y observations vary from the nonlinear (curved) function that is used to predict Y.

Nonlinear regression uses logarithmic functions, trigonometric functions, exponential functions, power functions, Lorenz curves, Gaussian functions, and other fitting methods.

KEY

- Both linear and nonlinear regression predict Y responses from an X variable (or variables).
- Nonlinear regression is a curved function of an X variable (or variables) that is used to predict a Y variable
- Nonlinear regression can show a prediction of population growth over time.

Example of Nonlinear Regression

One example of how nonlinear regression can be used is to predict population growth over time. A scatterplot of changing population data over time shows that there seems to be a relationship between time and population growth, but that it is a nonlinear relationship, requiring the use of a nonlinear regression model. A logistic population growth model can provide estimates of the population for periods that were not measured, and predictions of future population growth.

Independent and dependent variables used in nonlinear regression should be quantitative. Categorical variables, like region of residence or religion, should be coded as binary variables or other types of quantitative variables.

In order to obtain accurate results from the nonlinear regression model, you should make sure the function you specify describes the relationship between the independent and dependent variables accurately. Good starting values are also necessary. Poor starting values may result in a model that fails to converge, or a solution that is only optimal locally, rather than globally, even if you've specified the right functional form for the model.

4. Stepwise Regression Modeling

While the other items we have talked about until now are specific types of models, stepwise regression is more of a technique. If a model involves many potential inputs, the analyst may start with the most directly correlated input variable to build a model. Once that is accomplished, the next step is to make the model more accurate.

To do that, additional input variables can be added to the model one at a time in order of significance of the results. Using our email marketing example, the initial

model may be based on just the number of emails sent. Then we would add something like the average age of the email recipient. After that, we would add the average number of emails each recipient has received from us. Each additional variable would add a small amount of additional accuracy to the model.

This process is a stepwise modeling approach to getting the most accurate model. Alternatively, the analyst may start with a larger set of input variables and then incrementally remove the least significant in order to get to a desired model.

Multivariate analysis

Multivariate means involving multiple dependent variables resulting in one outcome. This explains that the majority of the problems in the real world are Multivariate. For example, we cannot predict the weather of any year based on the season. There are multiple factors like pollution, humidity, precipitation, etc.

Multivariate analysis (MVA) is a Statistical procedure for analysis of data involving more than one type of measurement or observation. It may also mean solving problems where more than one dependent variable is analyzed simultaneously with other variables.

Advantages and Disadvantages of Multivariate Analysis

Advantages

- The main advantage of multivariate analysis is that since it considers more than one factor of independent variables that influence the variability of dependent variables, the conclusion drawn is more accurate.
- The conclusions are more realistic and nearer to the real-life situation.

Disadvantages

- The main disadvantage of MVA includes that it requires rather complex computations to arrive at a satisfactory conclusion.

- Many observations for a large number of variables need to be collected and tabulated; it is a rather time-consuming process.

Multivariate analysis technique

Multivariate analysis technique can be classified into two broad categories viz., This classification depends upon the question: are the involved variables dependent on each other or not?

If the answer is yes: We have **Dependence methods**.

If the answer is no: We have **Interdependence methods**.

Dependence technique: Dependence Techniques are types of multivariate analysis techniques that are used when one or more of the variables can be identified as dependent variables and the remaining variables can be identified as independent.

Dependence Technique

Multiple Regression

Conjoint analysis

Multiple Discriminant Analysis

Linear Probability Models

Multivariate Analysis of Variance and Covariance

Canonical Correlation Analysis

Structural Equation Modeling

Multiple Discriminant Analysis

The objective of discriminant analysis is to determine group membership of samples from a group of predictors by finding linear

combinations of the variables which maximize the differences between the variables being studied, to establish a model to sort objects into their appropriate populations with minimal error.

Discriminant analysis derives an equation as a linear combination of the independent variables that will discriminate best between the groups in the dependent variable. This linear combination is known as the discriminant function. The weights assigned to each independent variable are corrected for the interrelationships among all the variables. The weights are referred to as discriminant coefficients.

The discriminant equation:

$$F = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p + \varepsilon$$

where, F is a latent variable formed by the linear combination of the dependent variable, X_1, X_2, \dots, X_p is the p independent variable, ε is the error term and $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ is the discriminant coefficients.

Interdependence Technique

Interdependence techniques are a type of relationship that variables cannot be classified as either dependent or independent.

It aims to unravel relationships between variables and/or subjects without explicitly assuming specific distributions for the variables. The idea is to describe the patterns in the data without making (very) strong assumptions about the variables.

Interdependence Technique

Factor Analysis

Cluster Analysis

Multidimensional Scaling

Correspondence Analysis

The Objective of multivariate analysis

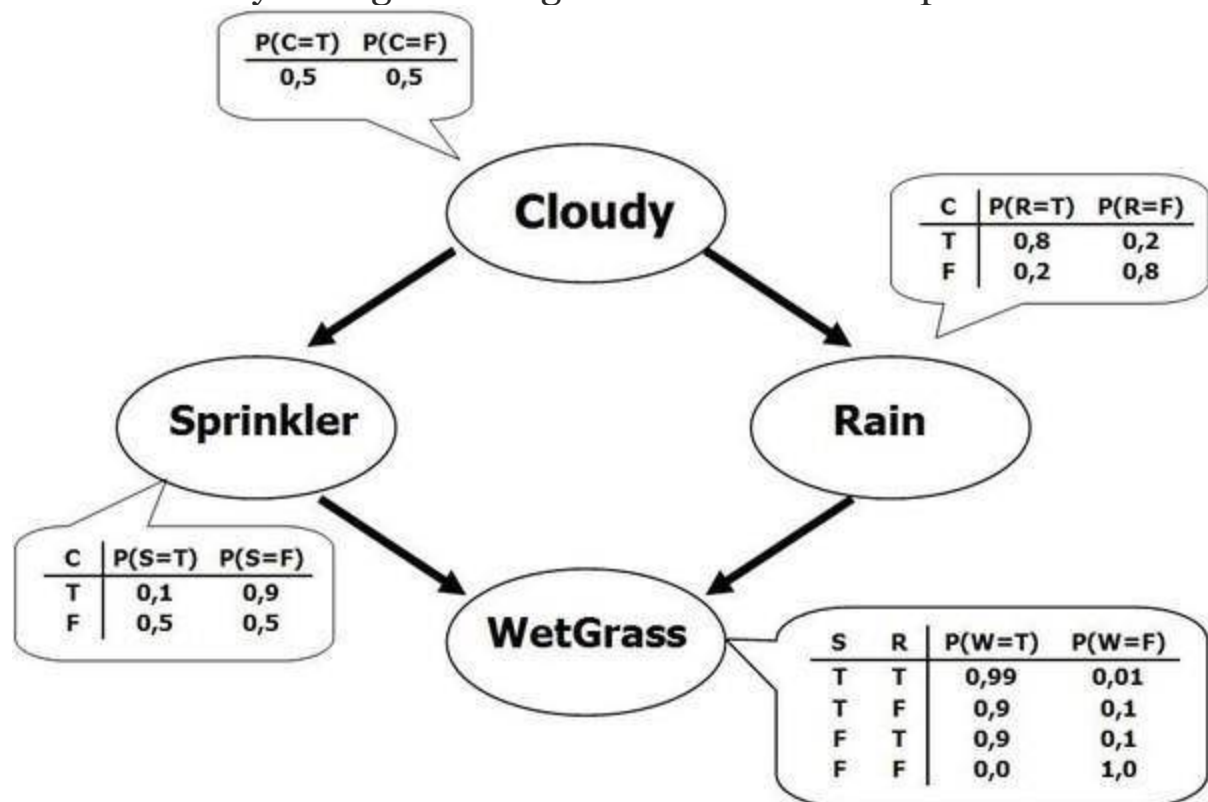
- (1) **Data reduction or structural simplification:** This helps data to get simplified as possible without sacrificing valuable information. This will make interpretation easier.
- (2) **Sorting and grouping:** When we have multiple variables, Groups of “similar” objects or variables are created, based upon measured characteristics.
- (3) **Investigation of dependence among variables:** The nature of the relationships among variables is of interest. Are all the variables mutually independent or are one or more variables dependent on the others?
- (4) **Prediction Relationships between variables:** must be determined for the purpose of predicting the values of one or more variables based on observations on the other variables.
- (5) **Hypothesis construction and testing.** Specific statistical hypotheses, formulated in terms of the parameters of multivariate populations, are tested. This may be done to validate assumptions or to reinforce prior convictions.

Bayesian Modeling

Bayesian networks are a type of probabilistic graphical model that uses Bayesian inference for probability computations. Bayesian networks aim to model conditional dependence, and therefore causation, by representing conditional dependence by edges in a directed graph. Through these relationships, one can efficiently conduct inference on the random variables in the graph through the use of factors.

The Bayesian Network

Using the relationships specified by our Bayesian network, we can obtain a compact, factorized representation of the joint probability distribution by taking advantage of conditional independence.



A Bayesian network is a **directed acyclic graph** in which each edge corresponds to a conditional dependency, and each node corresponds to a unique random variable. Formally, if an edge (A, B)

exists in the graph connecting random variables A and B, it means that $P(B|A)$ is a **factor** in the joint probability distribution, so we must know $P(B|A)$ for all values of B and A in order to conduct inference. In the above example, since Rain has an edge going into WetGrass, it means that $P(\text{WetGrass}|\text{Rain})$ will be a factor, whose probability values are specified next to the WetGrass node in a conditional probability table.

Bayesian networks satisfy the **local Markov property**, which states that a node is conditionally independent of its non-descendants given its parents. In the above example, this means that $P(\text{Sprinkler}|\text{Cloudy}, \text{Rain}) = P(\text{Sprinkler}|\text{Cloudy})$ since Sprinkler is conditionally independent of its non-descendant, Rain, given Cloudy. This property allows us to simplify the joint distribution, obtained in the previous section using the chain rule, to a smaller form. After simplification, the joint distribution for a Bayesian network is equal to the product of $P(\text{node}|\text{parents}(\text{node}))$ for all nodes,

Inference

Inference over a Bayesian network can come in two forms.

The first is simply evaluating the joint probability of a particular assignment of values for each variable (or a subset) in the network. For this, we already have a factorized form of the joint distribution, so we simply evaluate that product using the provided conditional probabilities. If we only care about a subset of variables, we will need to marginalize out the ones we are not interested in. In many

cases, this may result in underflow, so it is common to take the logarithm of that product, which is equivalent to adding up the individual logarithms of each term in the product.

The second, more interesting inference task, is to find $P(x|e)$, or, to find the probability of some assignment of a subset of the variables (x) given assignments of other variables (our evidence, e). In the above example, an example of this could be to find $P(\text{Sprinkler}, \text{WetGrass} \mid \text{Cloudy})$, where $\{\text{Sprinkler}, \text{WetGrass}\}$ is our x , and $\{\text{Cloudy}\}$ is our e . In order to calculate this, we use the fact that $P(x|e) = P(x, e) / P(e) = \alpha P(x, e)$, where α is a normalization constant that we will calculate at the end such that $P(x|e) + P(\neg x \mid e) = 1$. In order to calculate $P(x, e)$, we must marginalize the joint probability distribution over the variables that do not appear in x or e , which we will denote as Y .

$$P(x|e) = \alpha \sum_{\forall y \in Y} P(x, e, Y)$$

For the given example, we can calculate $P(\text{Sprinkler}, \text{WetGrass} \mid \text{Cloudy})$ as follows : $P(\text{Sprinkler}, \text{WetGrass} \mid \text{Cloudy}) = \text{Rain} \alpha \sum P(\text{WetGrass}|\text{Sprinkler}, \text{Rain})P(\text{Sprinkler}|\text{Cloudy})P(\text{Rain}|\text{Cloudy}) P(\text{Cloudy}) = \alpha P(\text{WetGrass}|\text{Sprinkler}, \text{Rain})P(\text{Sprinkler}|\text{Cloudy})P(\text{Rain}|\text{Cloudy}) P(\text{Cloudy}) + \alpha P(\text{WetGrass}|\text{Sprinkler}, \neg \text{Rain})P(\text{Sprinkler}|\text{Cloudy})P(\neg \text{Rain}|\text{Cloudy}) P(\text{Cloudy})$

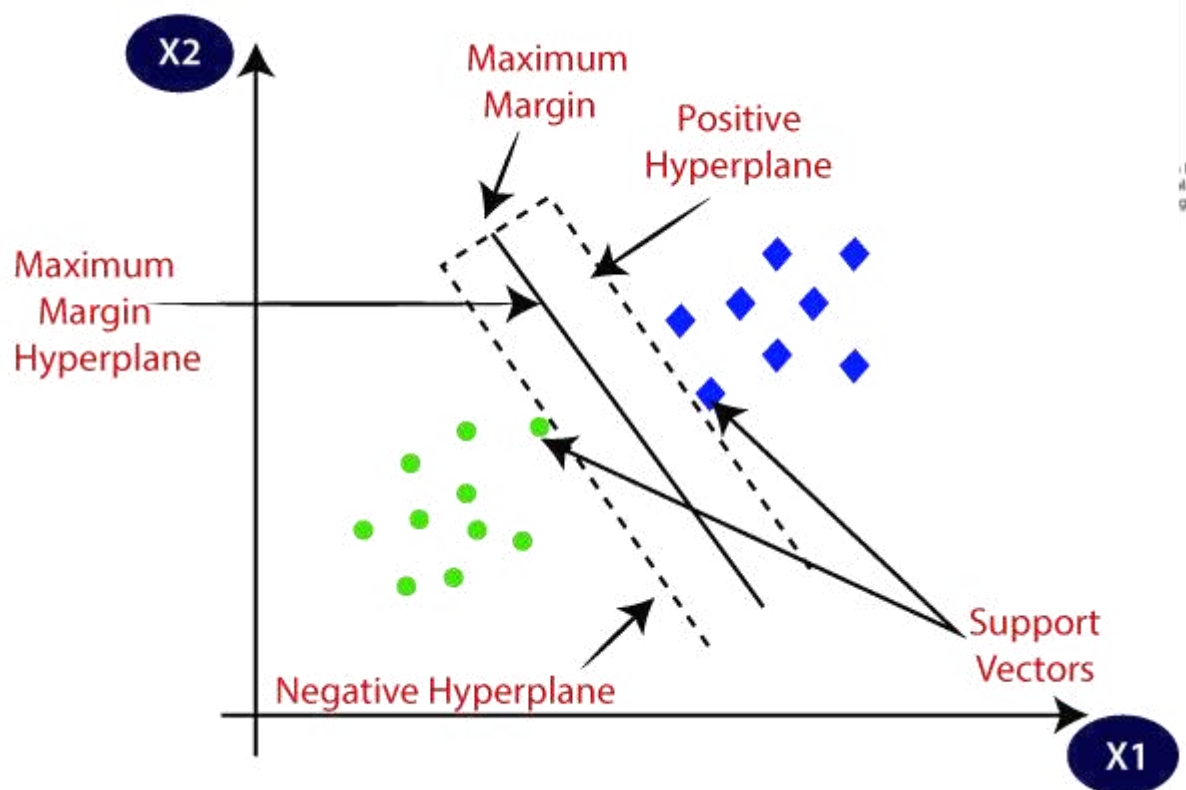
Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

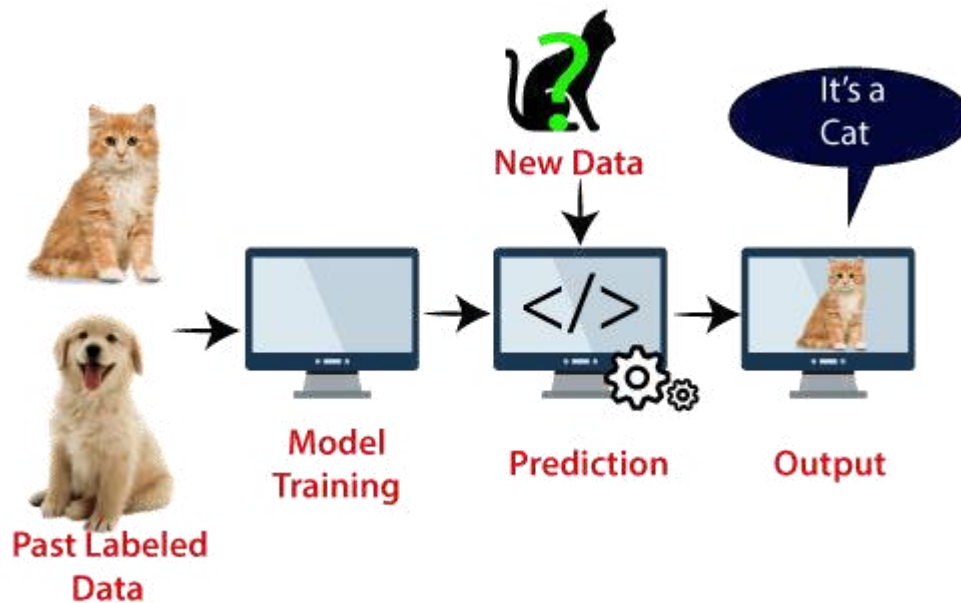
The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n -dimensional space into classes so that we can easily put the new data

point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



Example: SVM can be understood with the example that we have used in the KNN classifier. Suppose we see a strange cat that also has some features of dogs, so if we want a model that can accurately identify whether it is a cat or dog, so such a model can be created by using the SVM algorithm. We will first train our model with lots of images of cats and dogs so that it can learn about different features of cats and dogs, and then we test it with this strange creature. So as support vector creates a decision boundary between these two data (cat and dog) and choose extreme cases (support vectors), it will see the extreme case of cat and dog. On the basis of the support vectors, it will classify it as a cat. Consider the below diagram:



SVM algorithm can be used for **Face detection, image classification, text categorization**, etc.

Types of SVM

SVM can be of two types:

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

Hyperplane and Support Vectors in the SVM algorithm:

Hyperplane: There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

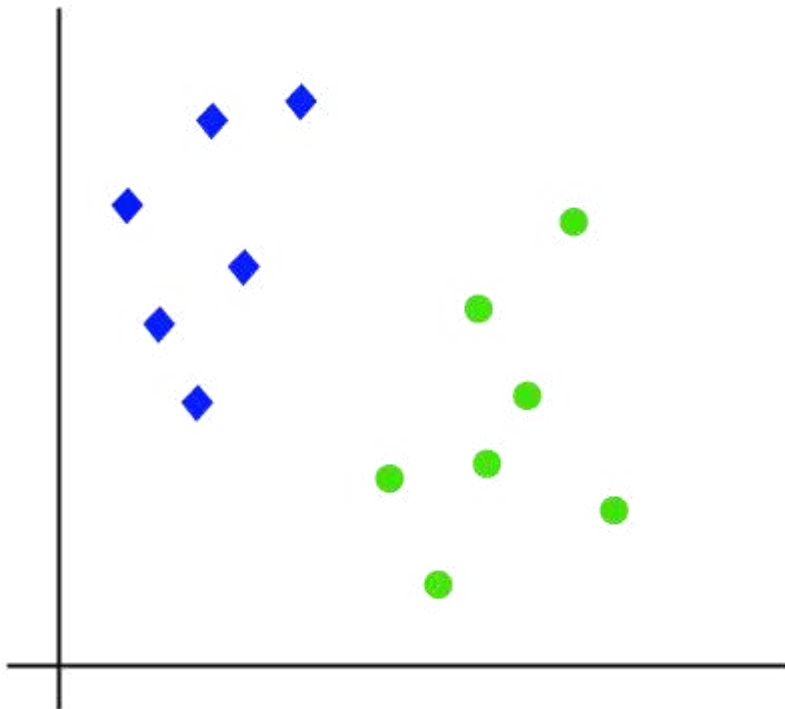
We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

Support Vectors:

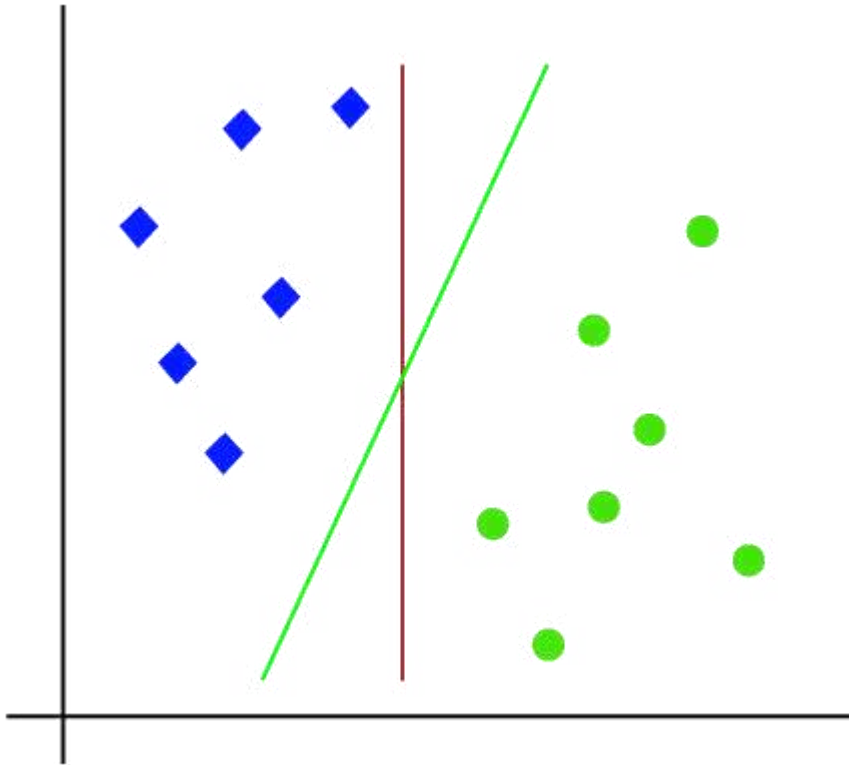
The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

How does SVM works?

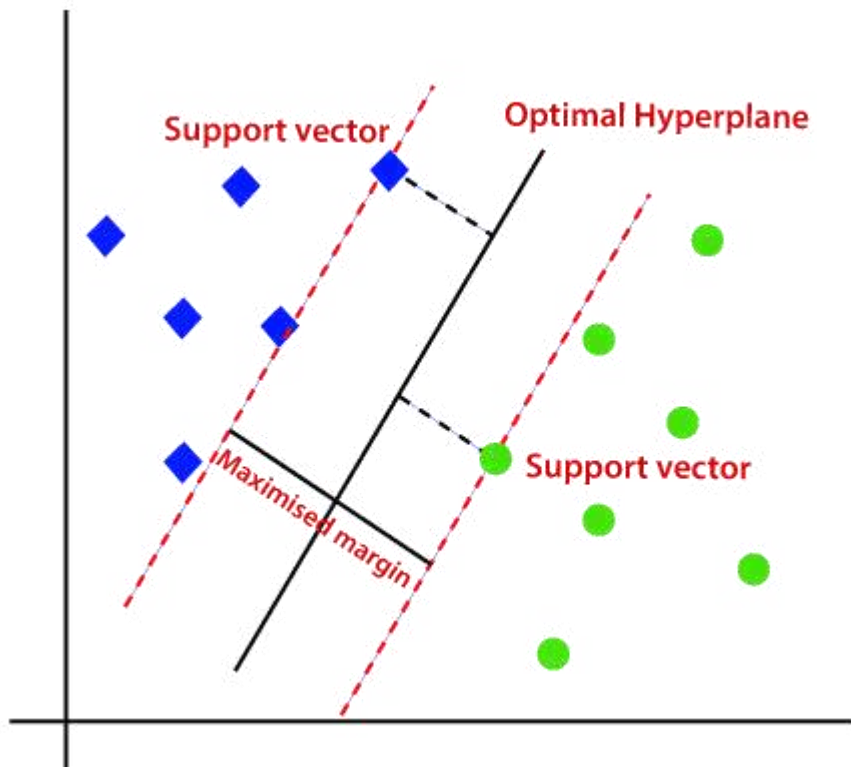
The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x_1 and x_2 . We want a classifier that can classify the pair(x_1 , x_2) of coordinates in either green or blue. Consider the below image:



So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:

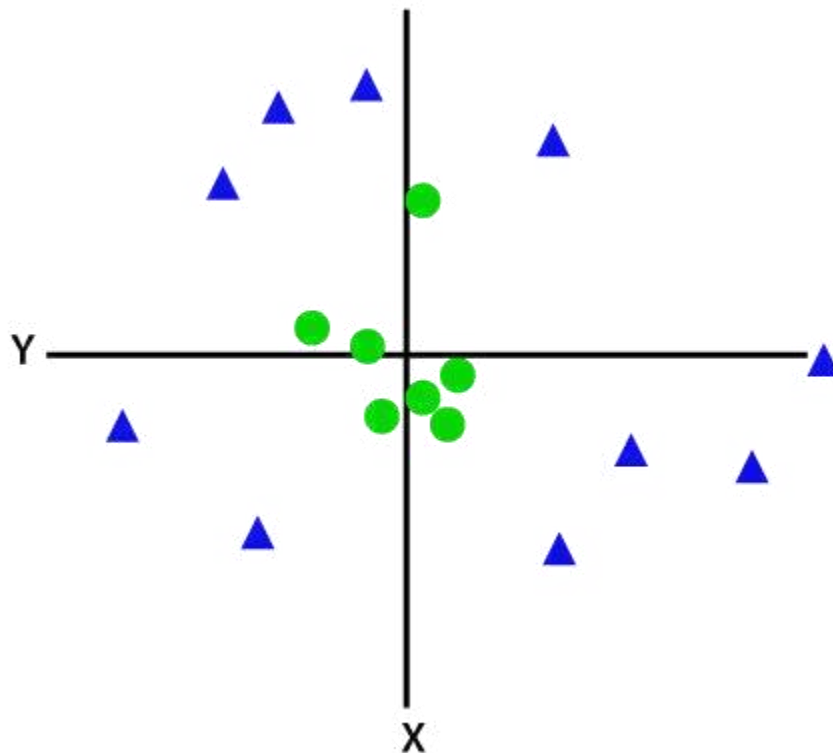


Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a **hyperplane**. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as **margin**. And the goal of SVM is to maximize this margin. The **hyperplane** with maximum margin is called the **optimal hyperplane**.



Non-Linear SVM:

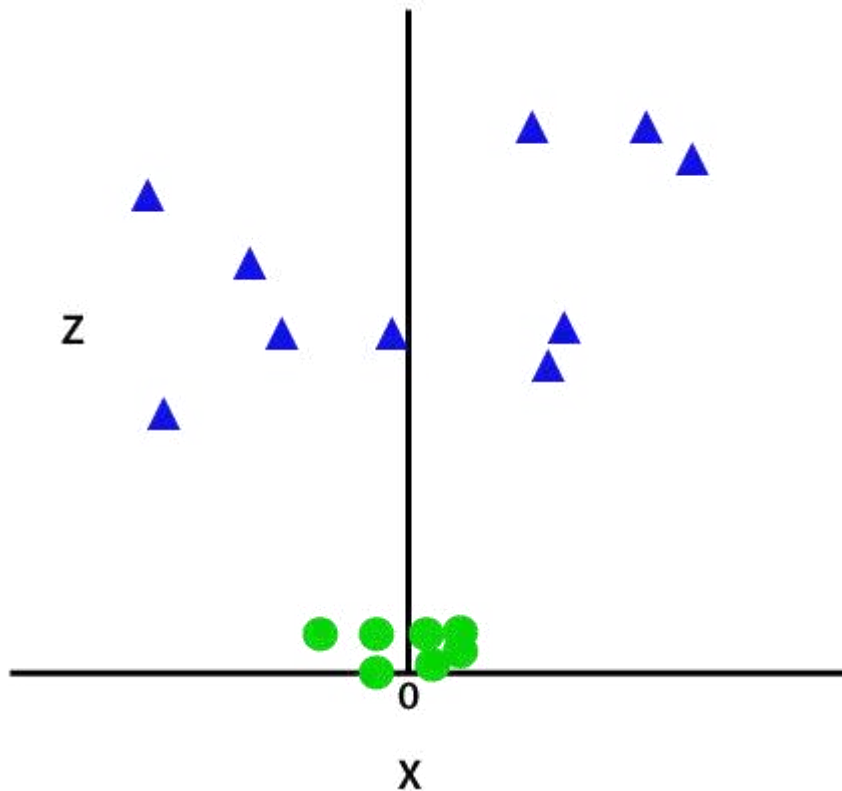
If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:



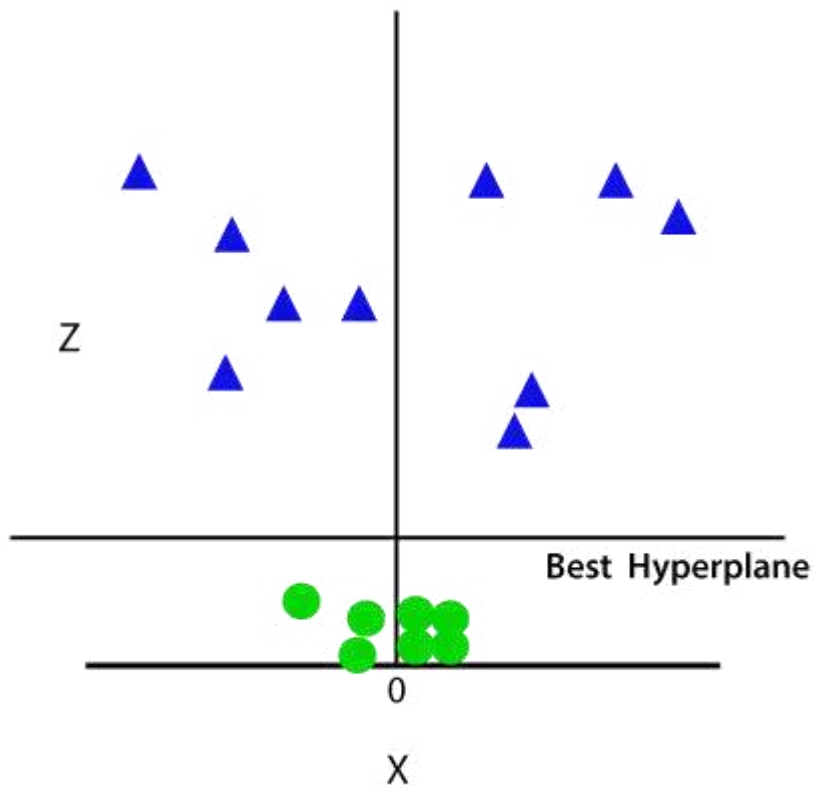
So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y , so for non-linear data, we will add a third dimension z . It can be calculated as:

$$z = x^2 + y^2$$

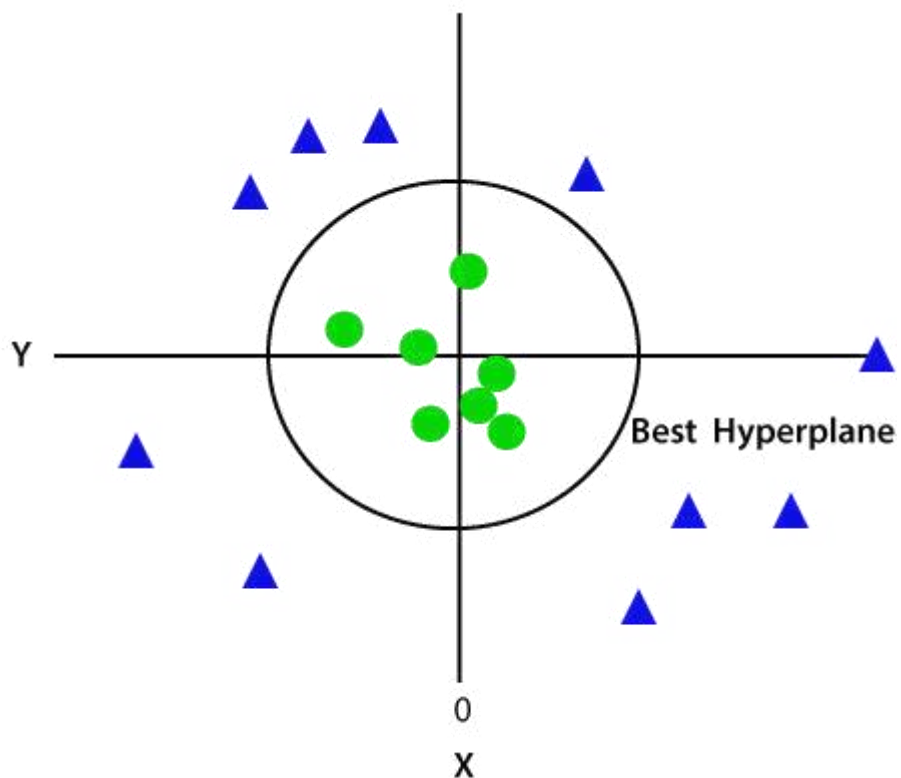
By adding the third dimension, the sample space will become as below image:



So now, SVM will divide the datasets into classes in the following way. Consider the below image:



Since we are in 3-d Space, hence it is looking like a plane parallel to the x-axis. If we convert it in 2d space with $z=1$, then it will become as:



Hence we get a circumference of radius 1 in case of non-linear data.

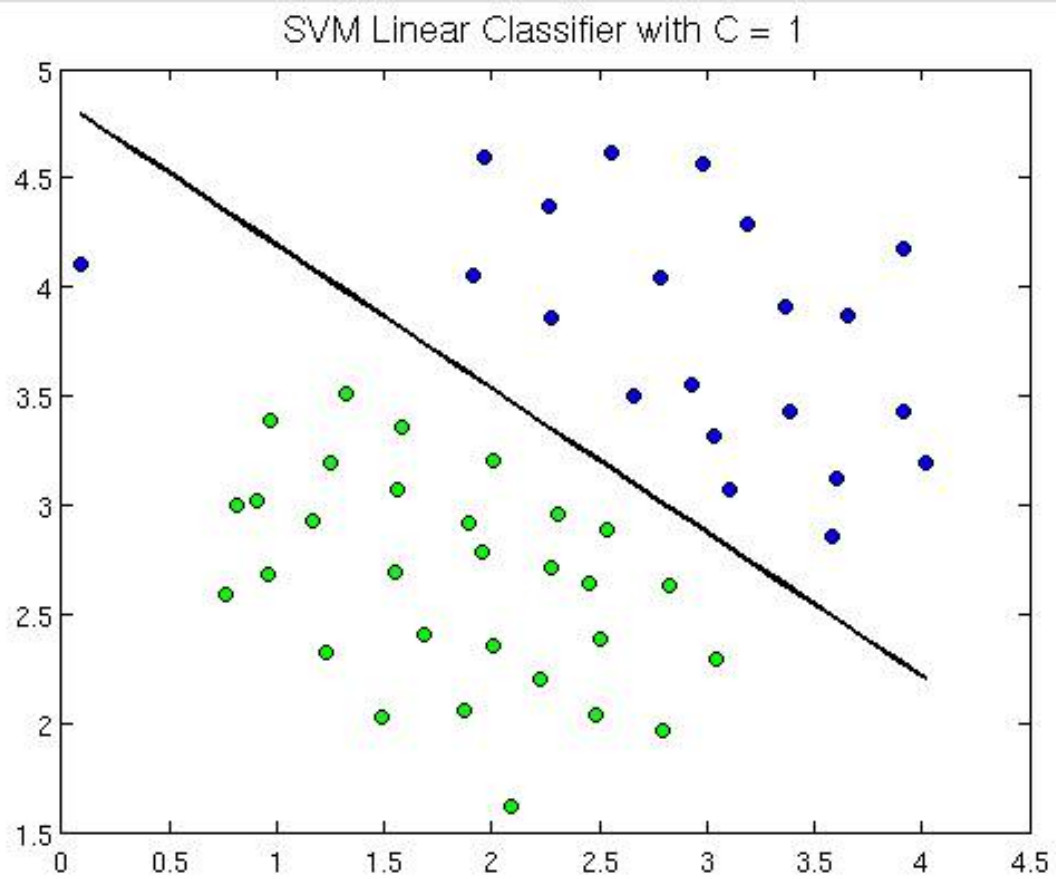
Introduction to Kernel Methods

Kernels or kernel methods (also called Kernel functions) are sets of different types of algorithms that are being used for pattern analysis. They are used to solve a non-linear problem by using a linear classifier. Kernel Methods are employed in SVM (Support Vector Machines) which are used in classification and regression problems. The SVM uses what is called a “Kernel Trick” where the data is transformed and an optimal boundary is found for the possible outputs.

The Need for Kernel Method and its Working

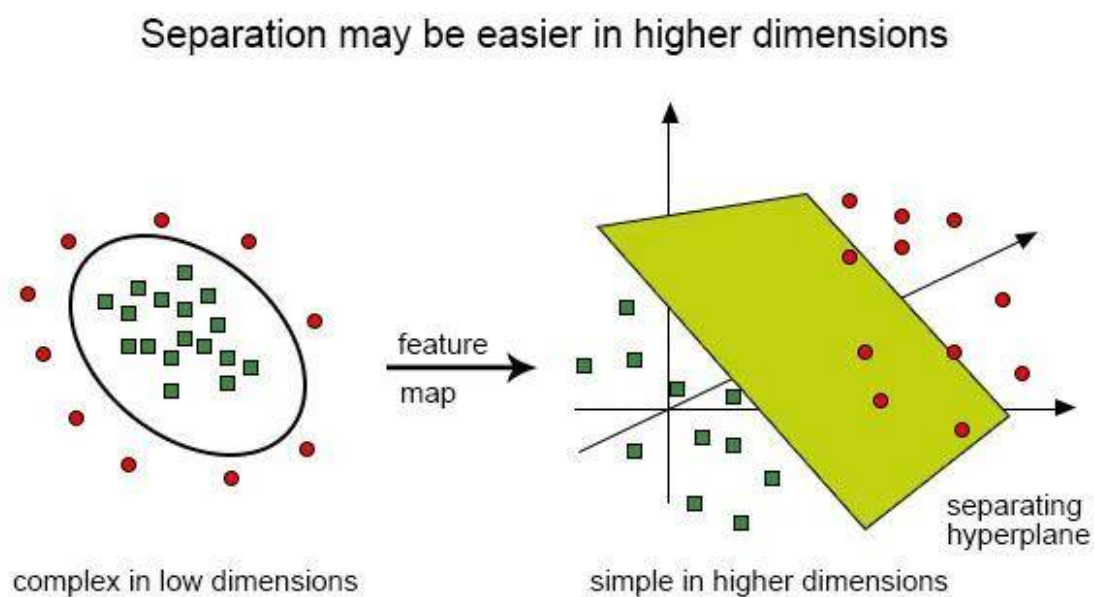
Before we get into the working of the Kernel Methods, it is more important to understand support vector machines or the SVMs because kernels are implemented in SVM models. So, Support Vector Machines are supervised [machine learning algorithms](#) that are used in classification and regression problems such as classifying an apple to class fruit while classifying a Lion to the class animal.

To demonstrate, below is what support vector machines look like:



Here we can see a hyperplane which is separating green dots from the blue ones. A hyperplane is one dimension less than the ambient plane. E.g. in the above figure, we have 2 dimension which represents the ambient space but the line which divides or classifies the space is one dimension less than the ambient space and is called hyperplane.

But what if we have input like this:



It is very difficult to solve this classification using a linear classifier as there is no good linear line that should be able to classify the red and the green dots as the points are randomly distributed. Here comes the use of kernel function which takes the points to higher dimensions, solves the problem

over there and returns the output. Think of this in this way, we can see that the green dots are enclosed in some perimeter area while the red one lies outside it, likewise, there could be other scenarios where green dots might be distributed in a trapezoid-shaped area.

So what we do is to convert the two-dimensional plane which was first classified by one-dimensional hyperplane (“or a straight line”) to the three-dimensional area and here our classifier i.e. hyperplane will not be a straight line but a two-dimensional plane which will cut the area.

In order to get a mathematical understanding of kernel, let us understand the Lili Jiang’s equation of kernel which is:

$K(x, y) = \langle f(x), f(y) \rangle$ where,

K is the kernel function,

X and Y are the dimensional inputs,

f is the map from n-dimensional to m-dimensional space and,

$\langle x, y \rangle$ is the dot product.

Illustration with the help of an example.

Let us say that we have two points, $x = (2, 3, 4)$ and $y = (3, 4, 5)$

As we have seen, $K(x, y) = \langle f(x), f(y) \rangle$.

Let us first calculate $\langle f(x), f(y) \rangle$

$$f(x) = (x_1x_1, x_1x_2, x_1x_3, x_2x_1, x_2x_2, x_2x_3, x_3x_1, x_3x_2, x_3x_3)$$

$$f(y) = (y_1y_1, y_1y_2, y_1y_3, y_2y_1, y_2y_2, y_2y_3, y_3y_1, y_3y_2, y_3y_3)$$

so,

$$f(2, 3, 4) = (4, 6, 8, 6, 9, 12, 8, 12, 16) \text{ and}$$

$$f(3, 4, 5) = (9, 12, 15, 12, 16, 20, 15, 20, 25)$$

so the dot product,

$$f(x) \cdot f(y) = f(2,3,4) \cdot f(3,4,5) =$$

$$(36 + 72 + 120 + 72 + 144 + 240 + 120 + 240 + 400) =$$

$$1444$$

And,

$$K(x, y) = (2 \cdot 3 + 3 \cdot 4 + 4 \cdot 5)^2 = (6 + 12 + 20)^2 = 38^2 = 1444.$$

This as we find out, $f(x) \cdot f(y)$ and $K(x, y)$ give us the same result, but the former method required a lot of calculations (because of projecting 3 dimensions into 9 dimensions) while using the kernel, it was much easier.

Types of Kernel and methods in SVM

Let us see some of the kernel function or the types that are being used in SVM:

1. *Liner Kernel*

Let us say that we have two vectors with name x_1 and Y_1 , then the linear kernel is defined by the dot product of these two vectors:

$$K(x_1, x_2) = x_1 \cdot x_2$$

2. *Polynomial Kernel*

A polynomial kernel is defined by the following equation:

$$K(x_1, x_2) = (x_1 \cdot x_2 + 1)^d,$$

Where,

d is the degree of the polynomial and x_1 and x_2 are vectors

3. *Gaussian Kernel*

This kernel is an example of a radial basis function kernel. Below is the equation for this:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$$

The given sigma plays a very important role in the performance of the Gaussian kernel and should neither be overestimated and nor be underestimated, it should be carefully tuned according to the problem.

4. Exponential Kernel

This is in close relation with the previous kernel i.e. the Gaussian kernel with the only difference is – the square of the norm is removed.

The function of the exponential function is:

$$k(x, y) = \exp \left(-\frac{\|x - y\|}{2\sigma^2} \right)$$

This is also a radial basis kernel function.

5. Laplacian Kernel

This type of kernel is less prone for changes and is totally equal to previously discussed exponential function kernel, the equation of Laplacian kernel is given as:

$$k(x, y) = \exp \left(-\frac{\|x - y\|}{\sigma} \right)$$

6. Hyperbolic or the Sigmoid Kernel

This kernel is used in [neural network](#) areas of machine learning. The activation function for the sigmoid kernel is the bipolar sigmoid function.

The equation for the hyperbolic kernel function is:

$$k(x, y) = \tanh(\alpha x^T y + c)$$

This kernel is very much used and popular among support vector machines.

7. Anova radial basis kernel

This kernel is known to perform very well in multidimensional regression problems just like the Gaussian and Laplacian kernels. This also comes under the category of radial basis kernel.

The equation for Anova kernel is :

$$k(x, y) = \sum_{k=1}^n \exp(-\sigma(x^k - y^k)^2)^d$$

There are a lot more types of Kernel Method and we have discussed the mostly used kernels. It purely depends on the type of problem which will decide the kernel function to be used.

What is time series analysis?

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time.

What sets time series data apart from other data is that the analysis can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data.

Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

Or

What is Time Series Analysis

Definition:

A time series is nothing but a sequence of various data points that occurred in a successive order for a given period of time

Objectives:

- To understand how time series works, what factors are affecting a certain variable(s) at different points of time.
- Time series analysis will provide the consequences and insights of features of the given dataset that changes over time.
- Supporting to derive the predicting the future values of the time series variable.
- Assumptions: There is one and the only assumption that is “stationary”, which means that the origin of time, does not affect the properties of the process under the statistical factor.

How to analyze Time Series?

Quick steps here for your reference, anyway. Will see this in detail in this article later.

- Collecting the data and cleaning it
- Preparing Visualization with respect to time vs key feature
- Observing the stationarity of the series
- Developing charts to understand its nature.
- Model building – AR, MA, ARMA and ARIMA
- Extracting insights from prediction

Significance of Time Series and its types

TSA is the backbone for prediction and forecasting analysis, specific to the time-based problem statements.

- Analyzing the historical dataset and its patterns
- Understanding and matching the current situation with patterns derived from the previous stage.
- Understanding the factor or factors influencing certain variable(s) in different periods.







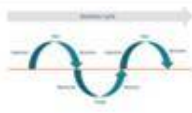
With help of “Time Series” we can prepare numerous time-based analyses and results.

- Forecasting
- Segmentation
- Classification

- Descriptive analysis`
- Intervention analysis

Components of Time Series Analysis

- Trend
- Seasonality
- Cyclical
- Irregularity
- **Trend:** In which there is no fixed interval and any divergence within the given dataset is a continuous timeline. The trend would be Negative or Positive or Null Trend
- **Seasonality:** In which regular or fixed interval shifts within the dataset in a continuous timeline. Would be bell curve or saw tooth
- **Cyclical:** In which there is no fixed interval, uncertainty in movement and its pattern
- **Irregularity:** Unexpected situations/events/scenarios and spikes in a short time span.

	Trend	Seasonality	Cyclical	Irregularity
Time	Fixed Time Interval	Fixed Time Interval	Not Fixed Time Interval	Not Fixed Time Interval
Duration	Long and Short Term	Short Term	Long and Short Term	Regular/Irregular
Visualization				
Nature - I	Gradual	Swings between Up or Down	Repeating Up and Down	Errored or High Fluctuation
Nature – II	Upward/Down Trend	Pattern repeatable	No fixed period	Short and Not repeatable
Prediction Capability	Predictable	Predictable	Challenging	Challenging
Market Model				Highly random/Unforeseen Events – along with white noise.

Designed by Author (Shanthababu)

What are the limitations of Time Series Analysis?

Time series has the below-mentioned limitations, we have to take care of those during our analysis,

- Similar to other models, the missing values are not supported by TSA
- The data points must be linear in their relationship.
- Data transformations are mandatory, so a little expensive.
- Models mostly work on Uni-variate data.

Time series analysis examples

Time series analysis is used for non-stationary data—things that are constantly fluctuating over time or are affected by time. Industries like finance, retail, and economics frequently use time series analysis because currency and sales are always changing. Stock market analysis is an excellent example of time series analysis in action, especially with automated trading algorithms. Likewise, time series analysis is ideal for forecasting weather changes, helping meteorologists predict everything from tomorrow's weather report to future years of climate change. Examples of time series analysis in action include:

- Weather data
- Rainfall measurements
- Temperature readings
- Heart rate monitoring (EKG)
- Brain monitoring (EEG)
- Quarterly sales
- Stock prices
- Automated stock trading
- Industry forecasts
- Interest rates

Rule Induction

Rule induction is a data mining process of deducing if-then rules from a data set. These symbolic decision rules explain an inherent relationship between the attributes and class labels in the data set. Many real-life experiences are based on intuitive rule induction. For example, we can proclaim a rule that states “if it is 8 a.m. on a weekday, then highway traffic will be heavy” and “if it is 8 p.m. on a Sunday, then the traffic will be light.” These rules are not necessarily right all the time. 8 a.m. weekday traffic may be light during a holiday season. But, in general, these rules hold true and are deduced from real-life experience based on our every day observations. Rule induction provides a powerful classification approach that can be easily understood by the general users. It is used in predictive analytics by classification of unknown data. Rule induction is also used to describe the patterns in the data. The easiest way to extract rules from a data set is from a decision tree that is developed on the same data set.

