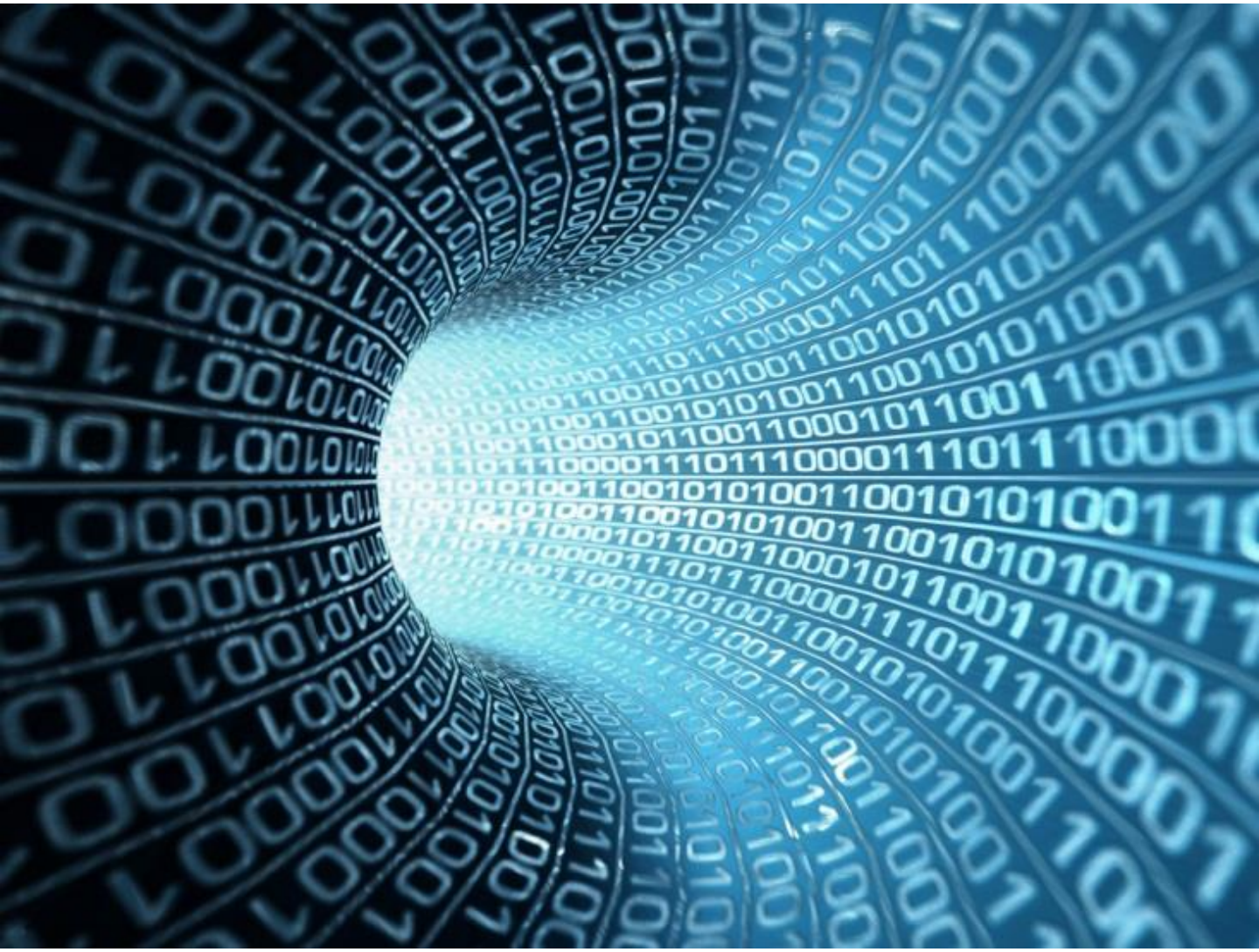


Newsletter Volume -9

Datamites
Global Institute for Data Science

Support Vector Machine Algorithm (SVM)

Support Vector Machine Algorithm (SVM) – Understanding Kernel Trick

For a better understanding, the blog has been split into two parts, the former gives conceptual clarity of Support Vector Machine (SVM) & Kernel Trick and the latter gives a mathematical explanation to the same. The blog also entails a complete modelling of the Support Vector Machine Algorithm using Python which will give us more confidence to embrace the algorithm and the concept.

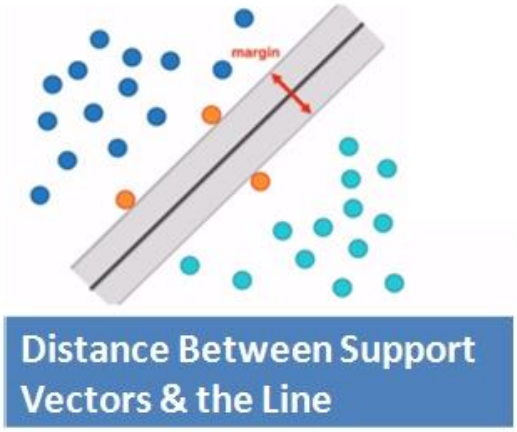
Understanding SVM and SVM Kernel Trick

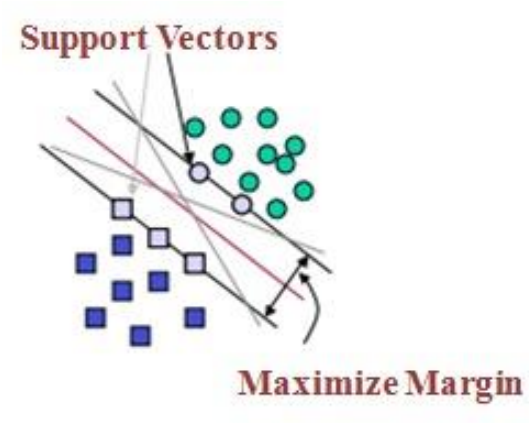
Before we decipher SVM Kernel Trick, let us first go over few basic concepts:

1. Support Vector Machine (SVM) and Support Vectors
2. Linearity, Non-Linearity, Dimensions and Planes
3. Kernel and Kernel methods

A **Support Vector Machine (SVM)** is a supervised machine learning algorithm which can be used for both classification and regression problems. Widely it is used for classification problem. SVM constructs a line or a hyper plane in a high or infinite dimensional space which is used for classification, regression or other tasks like outlier detection. SVM makes sure that the data is separated with the widest Margin.

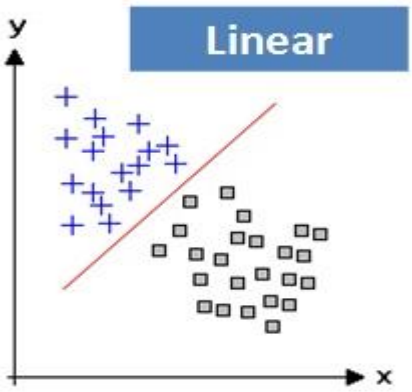
SVMs maximize the margin around the separating hyperplane.

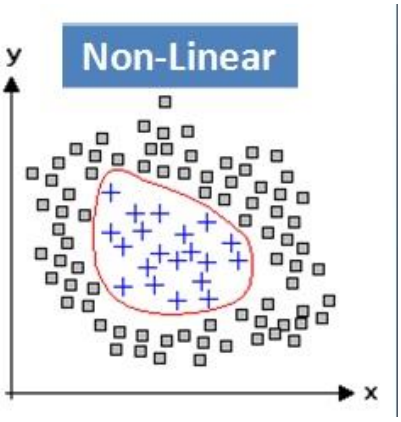




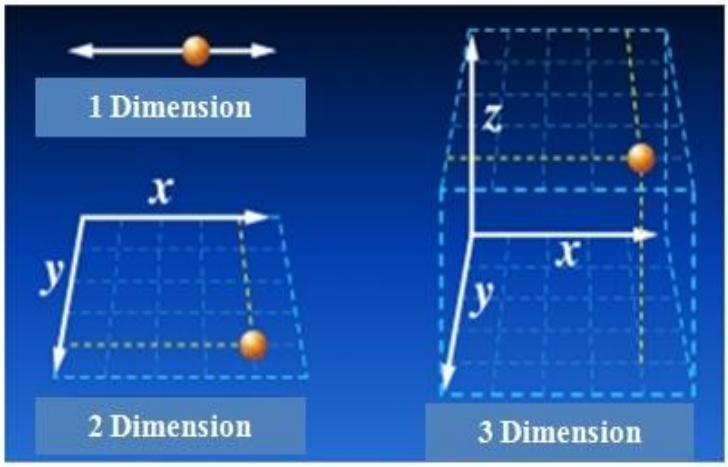
Support Vectors: Support vectors are the data points that lie farthest to the decision surface (or hyperplane). They are the data points which are most difficult to classify. They have a direct impact on the optimum location of the decision surface.

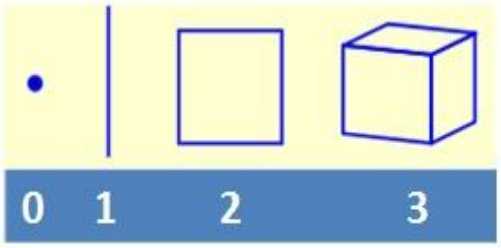
Linearity and Non-Linearity: On a Plane, a linear function is a function where the graph is a straight line. The line can go in any direction, but it's always a straight line. However, a non-linear function has a shape that is not a straight line.



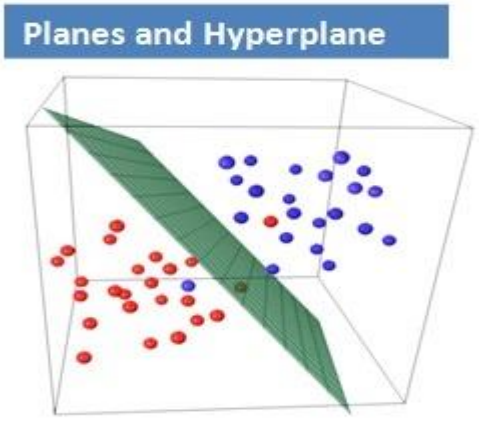


Dimensions: In simple terms, a dimension of something is a particular aspect of it. Examples: width, depth and height are dimensions. A line on a plane is one dimension, considering the edges a square has two dimensions and a cube has three dimensions.

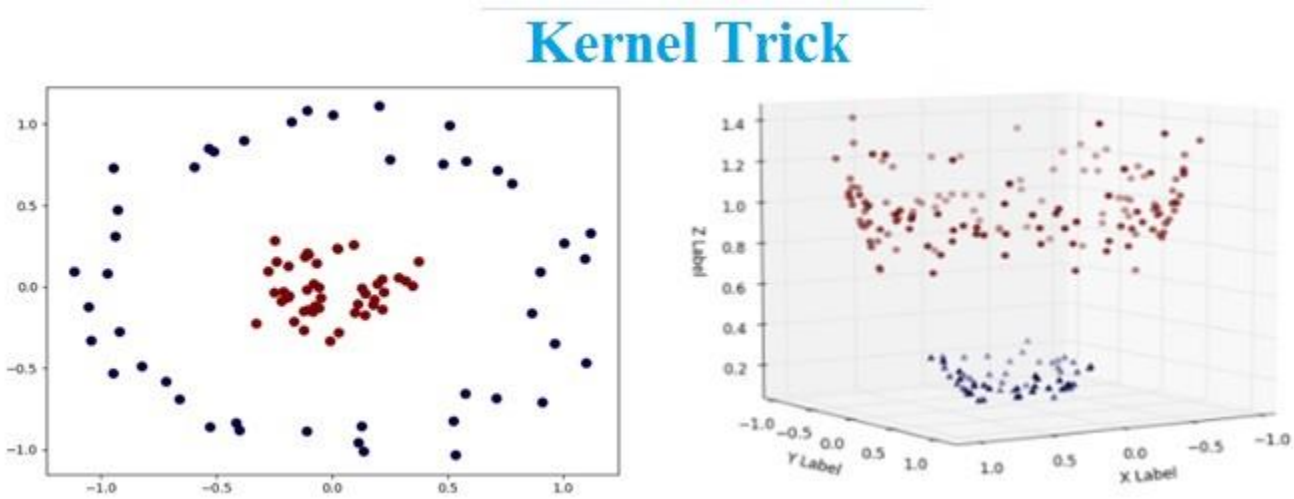




Planes and Hyperplane: In one dimension, a hyperplane is called a point. In two dimensions, it is a line. In three dimensions, it is a plane and in more dimensions we call it a hyperplane.

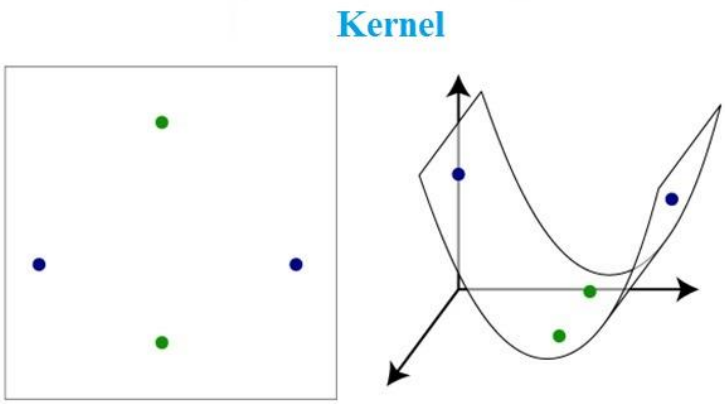


Kernel: A kernel is a method of placing a two dimensional plane into a higher dimensional space, so that it is curved in the higher dimensional space. (In simple terms, a kernel is a function from the low dimensional space into a higher dimensional space.)



After understanding the above it is now easier to relate and understand the Kernel Trick. We now know that SVM works better with two dimensional space which are linearly separable.

However, for a non-linear data SVM finds it difficult to classify the data. The easy solution here is to use the Kernel Trick. A Kernel Trick is a simple method where a Non Linear data is projected onto a higher dimension space so as to make it easier to classify the data where it could be linearly divided by a plane. This is mathematically achieved by Lagrangian formula using Lagrangian multipliers. (More details in the following mathematical section)



For Coding: https://gist.github.com/Sagar-DS/91ba1a3b5beb8bce78f541aa4861bc35

Advantages of Support Vector Machine

- Training of the model is relatively easy
- The model scales relatively well to high dimensional data
- SVM is a useful alternative to neural networks
- Trade-off amongst classifier complexity and error can be controlled explicitly
- It is useful for both Linearly Separable and Non-linearly Separable data
- Assured Optimality: The solution is guaranteed to be the global minimum due to the nature of Convex Optimization

Disadvantages of Support Vector Machine

- Picking right kernel and parameters can be computationally intensive
- In Natural Language Processing (NLP), a structured representation of text yields better performance. However, SVMs cannot accommodate such structures(word embedding)

SVM Applications

There are various applications of Kernel trick few of them are

- **Geostatistics:** It is a branch of statistics concentrating on spatial or spatiotemporal datasets. It was originally created to predict the probability distributions of ore grading at mining operations. Now it is applied in diverse disciplines including petroleum geology, hydrogeology, hydrology, meteorology, oceanography, geochemistry, geometallurgy, geography, forestry, environmental control, landscape ecology, soil science, and agriculture (specifically in precision farming).
- **Inverse distance weighting:** Type of deterministic method for multivariate interpolation with a known scattered set of points. The values assigned to unknown points are calculated with a weighted average of the values existing at the known points.
- **3D Reconstruction:** Process of capturing the shape and appearance of real objects.

- **Bioinformatics:** An interdisciplinary field that involves molecular biology, genetics, computer science, mathematics and statistics. Software tools and methods are developed to understand biological data better.

- **Chemoinformatics:** Application of computational and informational techniques over the field of chemistry to solve a wide range of problems.

- **Information Extraction:** Acronym as IE, It is a method of automated extraction or retrieval of structured information from an unstructured and semi-structured text documents, databases and websites.

- **Handwriting Recognition:** Acronym as HWR, It is the ability of a computer system to receive and interpret the handwritten input comprehensibly from different sources such as a letter paper documents, photographs, touch-screens and other devices.

Few Popular Kernels: The most tricky and demanding part of using SVM is to choose the right Kernel function because it's very challenging to visualize the data in n-dimensional space. Few popular kernels are:

- **Fisher Kernel:** It is a kernel function that analyses and measures the similarity of two objects. This is done on the basis of sets of measurements for each object and a statistical model.
- **Graph Kernel:** It is a kernel function that computes an inner product on graphs.
- **Polynomial Kernel:** It is a kernel commonly used with support vector machines (SVMs). It is also used with other kernelised models that symbolizes the similarity of vectors in a feature space over polynomials of the original variables, allowing learning of non-linear models.
- **Radial Basis Function Kernel (RBF):** It is a real-valued kernel function whose value depends only on the distance from the origin, or distance from some other point called a centre.

Upcoming Events

PyData is a group for users and developers of data analysis tools to share ideas and learn from each other. We gather to discuss how best to apply Python tools, as well as those using R and Julia, to meet the evolving challenges in data management, processing, analytics, and visualization.

PyData groups, events, and conferences aim to provide a venue for users across all the various domains of data analysis to share their experiences and their techniques. PyData is organized by NumFOCUS.org, a 501(c)3 non-profit in the United States.

EVENT NAME: PyData Bangalore Meetup #5

VENUE: Bangalore, India.

DATE: October 19, 2019

TIMINGS: 10.00 a.m.

VIEW MORE

Job Listings

Data Scientist

Foursis Technical Solutions

Years: 0 - 8 Location: Bangalore

Key Skills: SQL, MYSQL, SVN, Python, LINUX, Data Analysis.

Salary: Not Disclosed Posted On: 9 October

Apply

Data Scientist

6Sense

Years: 2 - 15 Location: Bangalore

Key Skills: Python, R, Machine learning, Deep learning, JAVA.

Salary: Not Disclosed Posted on: 9 October

Apply

Data Scientist

Amazon

Years: 0 - 7 Location: Bangalore

Key Skills: Experience in Python, R or other Programming Languages.

Salary: Not Disclosed Posted On: 6 October

Apply

Data Scientist

Cure . Fit

Years: 2- 5 Location: Bangalore

Key Skills: Proficiency in at least one Data Science Language (Python, Scala, R)

Salary: Not Disclosed Posted on: 3 October

Apply

Data Scientist

OTO Capital

Years: 0 - 5 Location: Bangalore

Key Skills: Experience in SQL, Big Data.

Salary: Not Disclosed Posted On: 8 October

Apply

Data Scientist

Technology For Efficiency

Years: 0 - 5 Location: Bangalore

Key Skills: Python, R, Machine learning, Deep learning, JAVA.

Salary: Not Disclosed Posted on: 7 October

Apply

Data Scientist

Western Digital

Years: 6 - 12 Location: Bangalore

Key Skills: Experience in SQL, Big Data.

Salary: Not Disclosed Posted On: 7 October

Apply

Data Scientist

Zycus

Years: 10 - 15 Location: Bangalore

Key Skills: Python, R, Machine learning, Deep learning, JAVA.

Salary: Not Disclosed Posted on: 5 October

Apply

Data Scientist

Bridgei2i Solutions Private Limited

Years: 0 - 5 Location: Bangalore

Key Skills: Experience in SQL, Big Data.

Salary: Not Disclosed Posted On: 7 October

Apply

Data Scientist

Pegasystems

Years: 0 - 5 Location: Bangalore

Key Skills: Python, R, Machine learning, Deep learning, JAVA.

Salary: Not Disclosed Posted on: 7 October

Apply

Data Scientist

SpeedLabs

Years: 0 - 5 Location: Bangalore

Key Skills: Experience in SQL, Big Data.

Salary: Not Disclosed Posted On: 9 October

Apply

Data Scientist

ANSR

Years: 0 - 5 Location: Bangalore

Key Skills: Python, R, Machine learning, Deep learning, JAVA.

Salary: Not Disclosed Posted on: 7 October

Apply

Senior Data Scientist

Reliance Industries Limited

Years: 0 - 9 Location: Bangalore

Key Skills: Experience in SQL, Big Data.

Salary: Not Disclosed Posted On: 6 October

Apply

Data Scientist

Think Bumbleblee

Years: 0 - 5 Location: Bangalore

Key Skills: Python, R, Machine learning, Deep learning, JAVA.

Salary: Not Disclosed Posted on: 5 October

Apply

Datamites

Global Institute for Data Science

:1800-313-3434

Email: info@datamites.com
www.datamites.com

#SwitchToDataScience