Python Packages for   
Data Science



Anish Mahapatra

Contents

[Introduction 4](#_Toc115638041)

[What is Support Vector Machines (SVM) in Machine Learning? 4](#_Toc115638042)

[Examples of SVM 5](#_Toc115638043)

[Why SVMs are used in Machine Learning 5](#_Toc115638044)

[Types of Support Vector Machines Algorithm 5](#_Toc115638045)

[Linear SVM 5](#_Toc115638046)

[Non-Linear SVM 5](#_Toc115638047)

[Hyperplane and Support Vectors in the SVM Algorithm 5](#_Toc115638048)

[How do we find the right hyperplane? 5](#_Toc115638049)

[How does SVM work in Machine Learning? 5](#_Toc115638050)

[Nonlinear data 5](#_Toc115638051)

[The Kernel Trick 5](#_Toc115638052)

[SVM Kernel Functions 5](#_Toc115638053)

[Simple SVM Classifier [Step-by-step] 5](#_Toc115638054)

[1.Create a new classifier 5](#_Toc115638055)

[2. Select how you want to classify your data 5](#_Toc115638056)

[3. Import your training data 5](#_Toc115638057)

[4. Define the tags for your SVM classifier 5](#_Toc115638058)

[5. Tag data to train your classifier 6](#_Toc115638059)

[6. Set your algorithm to SVM 6](#_Toc115638060)

[7. Test your classifier 6](#_Toc115638061)

[8. Integrate the topic classifier 6](#_Toc115638062)

[Applications of Support Vector Machine 6](#_Toc115638063)

[Advantages and Disadvantages of Support Vector Machine 6](#_Toc115638064)

[Conclusion 6](#_Toc115638065)

[Author 6](#_Toc115638066)

# Introduction

Welcome to the blog that is going to help you with Support Vector Machines. Machine Learning has been advancing over the years and so have the questions that are being asked in interviews. The focus has gone from the fundamental mathematics questions on simple linear regression and logisitic regression to more advanced machine learning algorithms. Interviewers tend to put forth more focus on models that you might not have seen as much, but are regularly used in the enterprise world to build and deploy models. One of the more popular models that is computationally acceptable and has great results is Support Vector Machines in machine learning. A model that is used for classification as well as regression, fundamental knowledge on support vector regression and classification should be known by all interviewees. In the following sections, we will go over the svm algorithm or svm machine learning to learn more about how to model works in depth, so that you are able to add support vector machine algorithm to your data science toolkit. In the following section, we will learn more about what is svm in machine learning.

# What is Support Vector Machines (SVM) in Machine Learning?

The svm model, or support vector machine model is a popular set of supervised learning models that is used for regression as well as classification analysis. It is a model that us based on the statstical learning framework and is known for being robust and effective in multiple use-cases. Based on a non-probabilistic binary linear classifier, support vector machine is used for seprating different classes with the help of various kernels that we will discuss in further sections. One of the main reasons that companies are leaning towards support vector mahine models as compares to other models is because support vector machines have significantly higher accuracy that can be leveraged, while using decreased computation from the system. One quick point to note here – SVM applications are generally implemented in the field of classification.

The questions on which kernel to choose while performing minimal computation is a big one, especially when we deal with larger datasets and this is done using something called the “kernel trick” and we will deep-dive into this topic in detail in a later section. Let us first get an intuition of support vector machines by looking at a few examples.

## Examples of SVM

In this section, we will look at a few svm examples. Let us forget all of this complex jargon that you might have read above and look at a solid example.

Q. What is the main goal of a classfication algorithm?

A. The main goal of a classification model in machine learning is to separate out different classes of points in an effective manner in a generalized manner. When doing this in a two dimensional (2-D) plane, it means drawing a straight line so that we are able to linearly separate out two classes of points in a manner that the future points also have a high probability of the points being separated out accurately.

Using the below support vector machine example, we will also introduce some new terminology.



[*Source*](https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm)

Let us understand sme simple terminology:

* **Hyperplane**: Similar to how a line can separate out points in a two dimensional space, a hyperplane is the plane that an separate out points in a n-dimensional space
* **Positive** **Hyperplane**: The dotted line that we see in the figure that is situated in the poitive region is called the positive hyperplane. The positive hyperplane passes through the first point in the positive space.
* **Negative** **Hyperplane**: The dotted line that we see in the figure that is situated in the negative region is called the negative hyperplane. The negative hyperplane passes through the first point in the negative space.
* **Hard** **Margin**: A hard margin indicates that the svm model is trying to work extremely well on the dataset and can cause overfitting. This is generally used in linearly separable data, generally only in linearly separable data.
* **Soft** **Margin**: The soft margin indicats that the model is flexible in terms of fitting the dataset and so, will not cause overfitting. This is used in most cases when the data is not linearly separable. It allows some extent of miscalssification to make the model fit better on the test dataset.
* **Maximum Margin Hyperplane**: The decision boundary (indiated in the above figure as a solid line) is the decision boundary based on which the points are bifurcated.

The idea behind selecting the decision boundary is that the larger the margin (difference between positive hyperplane and negative hyperplane), the lesser the generalization error as when we have smaller margins with the decision boundaries, it tends to lead to overfitting.

Besides this simple, yet effective example, support vector mahine is used to perform more complex use-cases such as categorization of text, classification of images and even face detection!

## Why SVMs are used in Machine Learning

The two main reasons why support vector machines are used in machine learning are:

* **Relatively high accuracy**: One of the main advantages of support vector macine is that as compared to more fundamental algorithms, it has a much higher relative accuracy. This means that when deploying the model in the real-world, we see better results from the machine learning models implemented.
* **Minimal Computation time**: Due to the “kernel trick”, the computation time of support vector machines is greatly reduced, which means that as data scientists, we are able to get beter results out in a reduced time, while utilizing lesses resources. This is a win-win, as we are able to get the better results, without affecting hardware utilization costs and even at a faster time!

In the next section, we will take a deep dive intor the types of support vector machine algorithms.

# Types of Support Vector Machines Algorithm

In this section, we will understand more about the types of SVMs, based on the kind of data that we use. This is more specific to classification as that is the primary use-ase for support vector machines.

## Linear SVM

The linear support vector machine algorithm is used when we have linearly separable data. In simple language, if we have a dataset that **can** be classified into two groups using a simple straight line, we call it linearly separable data, and the classifier used for this is known as Linear SVM classifier.

## Non-Linear SVM

The non-linear support vector machine algorithm is used when we have non-linearly separable data. In simple language, if we have a dataset that **cannot** be classified into two groups using a simple straight line, we call it non-linear separable data, and the classifier used for this is known as Non-Linear SVM classifier.

# Hyperplane and Support Vectors in the SVM Algorithm

## How do we find the right hyperplane?

# How does SVM work in Machine Learning?

## Nonlinear data

## The Kernel Trick

# SVM Kernel Functions

# Simple SVM Classifier [Step-by-step]

## 1.Create a new classifier

## 2. Select how you want to classify your data

## 3. Import your training data

## 4. Define the tags for your SVM classifier

## 5. Tag data to train your classifier

## 6. Set your algorithm to SVM

## 7. Test your classifier

## 8. Integrate the topic classifier

# Applications of Support Vector Machine

# Advantages and Disadvantages of Support Vector Machine

# Conclusion

# Author

**Anish Mahapatra**   
*Senior Data Scientist*   
LinkedIn: [https://www.linkedin.com/in/anishmahapatra](about:blank) ([LinkedIn](about:blank))   
Email: anishmahapatra01@gmail.com