Support Vector Machines

in Machine Learning



**Anish Mahapatra**

Contents

[Introduction 4](#_Toc115770504)

[What are Support Vector Machines (SVM) in Machine Learning? 4](#_Toc115770505)

[Examples of SVM 4](#_Toc115770506)

[Why SVMs are used in Machine Learning 7](#_Toc115770507)

[Types of Support Vector Machines Algorithm 7](#_Toc115770508)

[Linear SVM 7](#_Toc115770509)

[Non-Linear SVM 7](#_Toc115770510)

[Hyperplane and Support Vectors in the SVM Algorithm 7](#_Toc115770511)

[Hyperplane 8](#_Toc115770512)

[Support Vectors 8](#_Toc115770513)

[How do we find the right hyperplane? 8](#_Toc115770514)

[How does SVM work in Machine Learning? 10](#_Toc115770515)

[Linearly separable data 10](#_Toc115770516)

[Non-linear data 11](#_Toc115770517)

[The Kernel Trick 12](#_Toc115770518)

[SVM Kernel Functions 12](#_Toc115770519)

[Simple SVM Classifier [Step-by-step] 14](#_Toc115770520)

[Import the required libraries 14](#_Toc115770521)

[Import the required dataset 14](#_Toc115770522)

[Maximize the margin 16](#_Toc115770523)

[Fit the support vector machine classifier 16](#_Toc115770524)

[Decide the kernel type based on the data distribution type 17](#_Toc115770525)

[Applications of Support Vector Machine 18](#_Toc115770526)

[Advantages and Disadvantages of Support Vector Machine 19](#_Toc115770527)

[Advantages of SVM 19](#_Toc115770528)

[Disadvantages of SVM 20](#_Toc115770529)

[Conclusion 20](#_Toc115770530)

[FAQs 20](#_Toc115770531)

[What are support vector machines with examples? 20](#_Toc115770532)

[Which type of classifier is SVM? 21](#_Toc115770533)

[Which is better, SVM or neural network? 21](#_Toc115770534)

[What are the advantages of SVM? 21](#_Toc115770535)

[Can we use SVM for regression? 21](#_Toc115770536)

[Author 21](#_Toc115770537)

# 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 logistic 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 are 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 the support vector machine algorithm to your data science toolkit. In the following section, we will learn more about what svm is in machine learning. You can learn more in the [data science part time bootcamp](https://www.knowledgehut.com/data-science/data-science-bootcamp-training).

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

The SVM model or support vector machine model is a popular set of supervised learning models that are used for regression as well as classification analysis. It is a model that is based on the statistical learning framework and is known for being robust and effective in multiple use cases. Based on a non-probabilistic binary linear classifier, a support vector machine is used for separating 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 machine models as compared 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 question of which kernel to choose while performing minimal computation is a big one, especially when we deal with larger datasets. 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 classification 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.

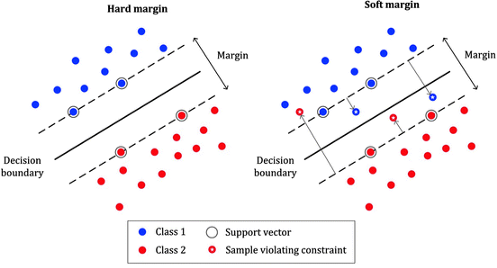
Using the below support vector machine example, we will also introduce some new terminology.



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

Let us understand some simple terminology:

* **Hyperplane**: Similar to how a line can separate out points in a two-dimensional space, a hyperplane is the plane that separates out points in an n-dimensional space
* **Positive** **Hyperplane**: The dotted line that we see in the figure that is situated in the positive 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 indicates 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 misclassification to make the model fit better on the test dataset.
* **Maximum Margin Hyperplane**: The decision boundary (indicated in the above figure as a solid line) is the decision boundary based on which the points are bifurcated.



*Figure: Soft Margin versus Hard Margin (*[*Source*](https://medium.com/swlh/the-support-vector-machine-basic-concept-a5106bd3cc5f#:~:text=SVM%20has%20a%20special%20way,finds%20the%20optimal%20separating%20hyperplane.)*)*

The idea behind selecting the decision boundary is that the larger the margin (difference between the 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, a support vector machine 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 a support vector machine 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 svm support vector machines is greatly reduced, which means that as data scientists, we are able to get better results in a reduced time while utilizing fewer resources. This is a win-win, as we are able to get better results without affecting hardware utilization costs and even at a faster time!

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

# Types of Support Vector Machines Algorithm

In this section, we will understand more about the types of SVM based on the kind of data that we use. This is more specific to classification as that is the primary use case 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 the 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 the Non-Linear SVM classifier.

# Hyperplane and Support Vectors in the SVM Algorithm

In this section, we will discuss more hyperplane and support vectors in svm.

## Hyperplane

When given a set of points, there can be multiple ways to separate the classes in an n-dimensional space. The way that SVM works, it transforms the lower dimensional data into higher dimensional data and then separates out the points. There are multiple ways to separate out the data, and these can be called decision boundaries. However, the main idea behind svm classification is to find the best possible decision boundary. The hyperplane is the optimal, generalized and best-fit boundary for the support vector machine classifier.

For instance, in a two-dimensional space, as we discussed in our example, the hyperplane will be s straight line, whereas, if the data exists in a three-dimensional space, then the hyperplane will exist in two dimensions. A good rule of thumb is that for an n-dimensional space, the hyperplane will generally have an n-1 dimension.

The aim is to create a hyperplane that has the highest possible margin so as to create a generalized model. This indicates that there will be a maximum distance between data points.

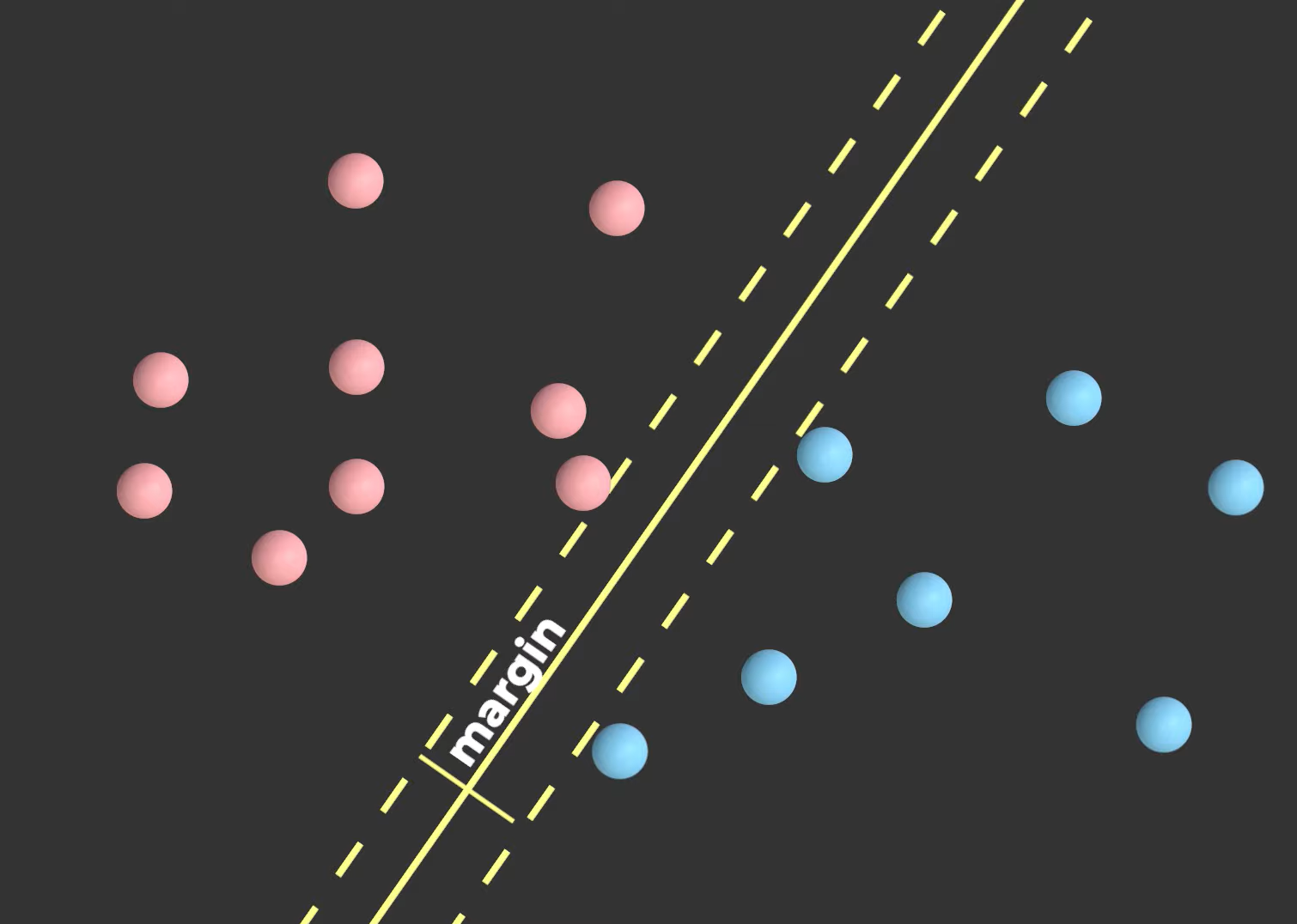
## Support Vectors

The term support vector indicates that we have supporting vectors to the main hyperplane. If we have the maximum distance between the support vectors, it is an indication of the best fit. So, support vectors are the vectors that pass through the closest points to the hyperplane and affect the overall position of the hyperplane.

## How do we find the right hyperplane?

Now, we come to a great question, how do we actually find the right hyperplane? Let us try to visualize and understand the two ways that we find the right hyperplane.

* **Maximize the margin between support vectors**The recommended way to find the right hyperplane is by maximizing the distance between the support vectors. Below, we visualize what this will look like in a two-dimensional space, this can also be done in an n-dimensional space, but it will be difficult for us to visualize.

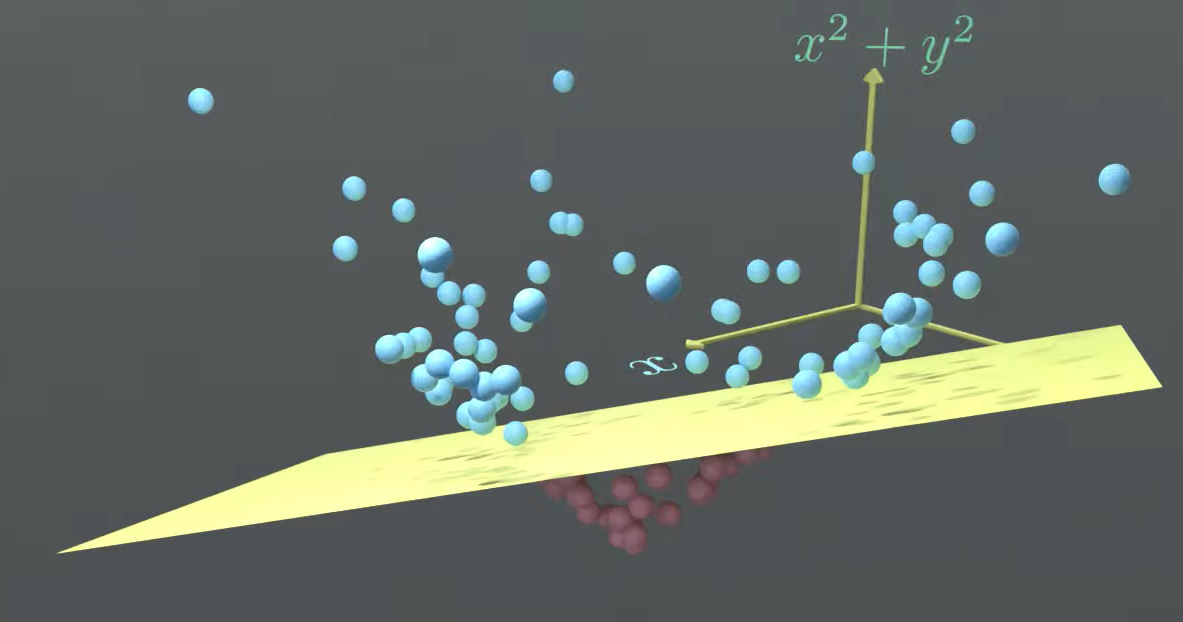


*Figure: Maximize the margin between the support vectors (*[*Source*](https://www.youtube.com/watch?v=_YPScrckx28)*)*

* **Transform lower dimensional data into higher dimensional data**When we transform lower dimensional data into higher dimensional data, with the help of new features created, it separates out the points in a higher dimension, and we can then pass a hyperplane with more efficiency to segregate out the data.



*Figure: Two-dimensional view of the data on the X and Y axes (*[*Source*](https://www.youtube.com/watch?v=_YPScrckx28)*)*

*******Figure: Projecting the data in a three-dimensional space (*[*Source*](https://www.youtube.com/watch?v=_YPScrckx28)*)*

This is done with the help of the following steps:

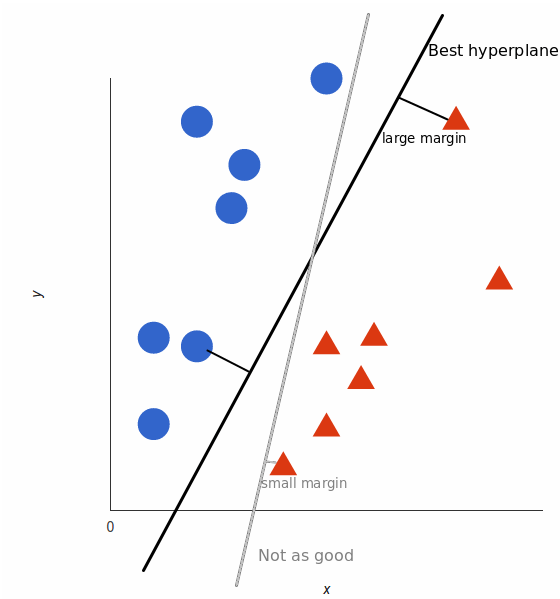
1. Augment the data with some non-linear features that are computed using the existing features
2. Find the separating hyperplane in the higher dimensional space
3. Project the points back to the original space

# How does SVM work in Machine Learning?

SVM works based on the principle of maximizing the distance between the support vectors. This ensures that we have the maximum margin possible between points, thus, giving us a generalized model. The aim of support vector machine classification is to maximize the margin between the support vectors. Let us go over the ways that SVM works in machine learning. You can learn more in the [knowledgehut data science part time bootcamp](https://www.knowledgehut.com/data-science/data-science-bootcamp-training).

## Linearly separable data

We use kernels in support vector machines. SVM kernels are functions based on which we can transform the data so that it is easier to fit a hyperplane to segregate the points better.

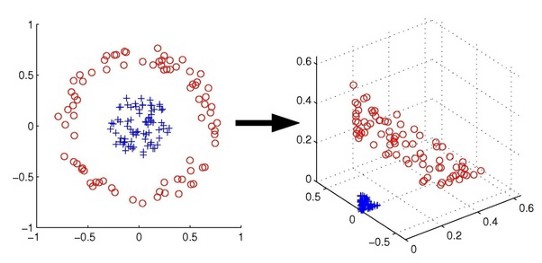


*Figure: Linearly separable data (*[*Source*](https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/)*)*

Linearly separable points generally consist of points that can be separated by a simple straight line. The line has to have the largest margin possible between the closest points to form a generalized svm model.

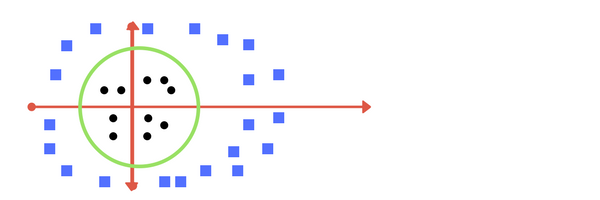
## Non-linear data

Non-linear data is data that cannot be separated via a simple straight line. We can separate out the classes by mapping the data into a higher dimensional space such that we are able to classify the points. Here, we use derived higher dimensional features from the dataset itself. For instance, with a dataset that is present on the X and Y axis, we will use features such as X2, Y2, and XY to make a higher dimensional model, project the data, make the hyperplane and then revert the data to its original space.



*Figure: A two-dimensional data depicted as three-dimensional data (*[*Source*](https://medium.com/swlh/the-support-vector-machine-basic-concept-a5106bd3cc5f#:~:text=SVM%20has%20a%20special%20way,finds%20the%20optimal%20separating%20hyperplane.)*)*

This is done using a clever trick that we will discuss in the next section. In the end, the figure will end up looking like the below figure, which separates the two classes in the same original space.



*Figure: Separating the points out by mapping the   
three-dimensional space to two-dimensional space (*[*Source*](https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72)*)*

## The Kernel Trick

The kernel trick is the “superpower” of support vector machines. Essentially, a support vector machine uses something called kernels, which is a function based on which the points can be segregated. The points that are no-linearly separable are projects to a higher dimensional space.

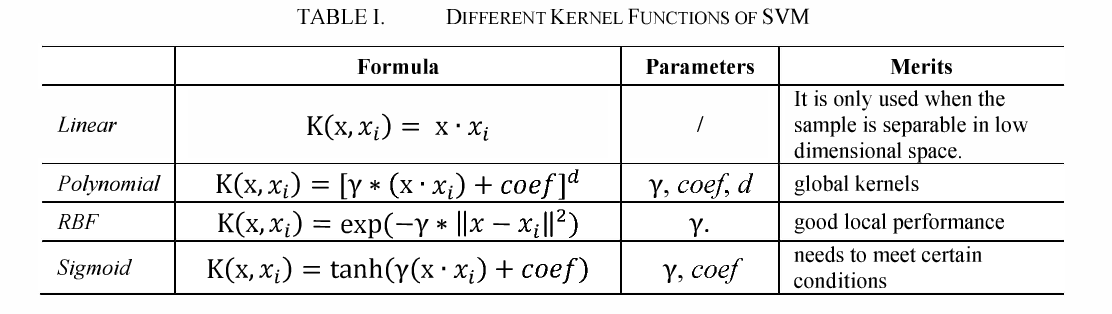
**Q. So, what is the “trick” here?**

A. SVM represents the non-linear data points in a fashion where the data points seem to be transformed and then find the hyperplane. However, in reality, the points actually remain the same, and they have not been transformed.

This trick is the reason that the seeming transformation of the points from a lower to a higher dimension is known as the kernel trick.

# SVM Kernel Functions

We have been talking about svm kernels for a while now. Let us briefly go over some of the important kernel functions that help transform the data to pass hyperplanes to segregate the data. All of the neat tricks we talk about are math; the transformations of the data are performed using linear algebra. We are going to go into a little bit of mathematics now, as this will help give you an intuition of the kernel.

*Table: Mathematical formulas of the kernel functions (*[*Source*](https://www.semanticscholar.org/paper/The-optimization-of-kernel-function-and-its-for-SVM-Ren-Hu/3a92a26a66efba1849fa95c900114b9d129467ac)*)*

* **Linear Kernel** **Function**  
  The linear kernel is primarily leveraged for linearly separable data. It is used for points that have a linear relationship.
* **Polynomial Kernel Function**   
  The polynomial kernel function is used by leveraging the dot product and transforming the data to an n-dimension. This helps represent the data with a higher dimension leveraging newly transformed data points.
* **RBF (Radial basis Function)**This is one of the most common and widely used functions as a kernel, which behaves similar to a weighted nearest neighbour model. It can transform the given data into infinite dimensions and then leverage the weighted nearest neighbour model to identify the observations that have the biggest influence on the new data point for the classification. The ‘Radial’ function in RBF can either be Laplace or Gaussian. We can decide this based on the ‘Gamma**’** hyperparameter**.**
* **Sigmoid function**The sigmoid function is found in use cases such as neural networks, where it is used as an activation function (Tanh). It is also known as the hyperbolic tangent function and has certain use cases where it is able to segregate the data better.

That was the support vector machine explained. We have now learnt about the various kernels that are used in support vector machine functions. Next, we will go over the svm classifier python code.

# Simple SVM Classifier [Step-by-step]

In this section, we will look over the svm implementation in Python. We will quickly go over an example of Python code to see support vector machines in action.

## Import the required libraries

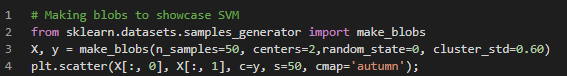
The support vector machine can be used from the SVC python library, which stands for support vector classifier. It is a supervised learning algorithm that is used to perform classification and can be found in the sci-kit learn. We can look at two use cases of a dataset that has a linear and non-linear distribution for this python showcase. We will make a blob to showcase the data in the next section.



*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

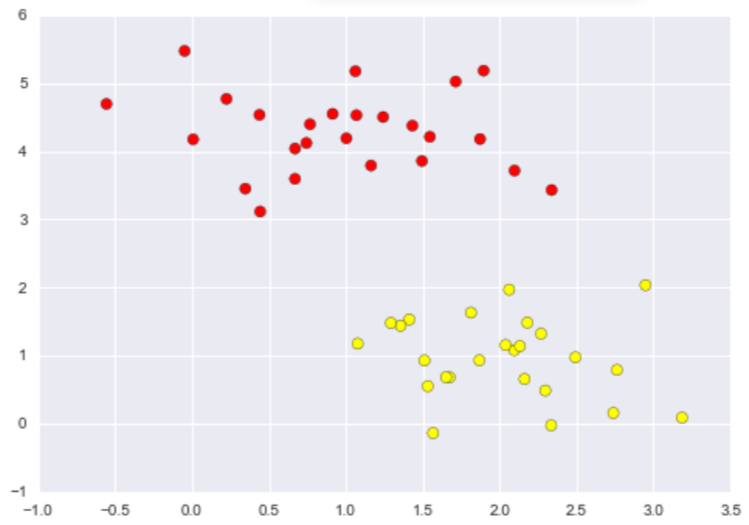
## Import the required dataset

Generally, in this case, we should import the required dataset, perform the necessary pre-processing steps, and then analyze and visualize the data. Here, in this case, we will generate two blobs to showcase the power of support vector machines using kernels that we have discussed in previous sections.



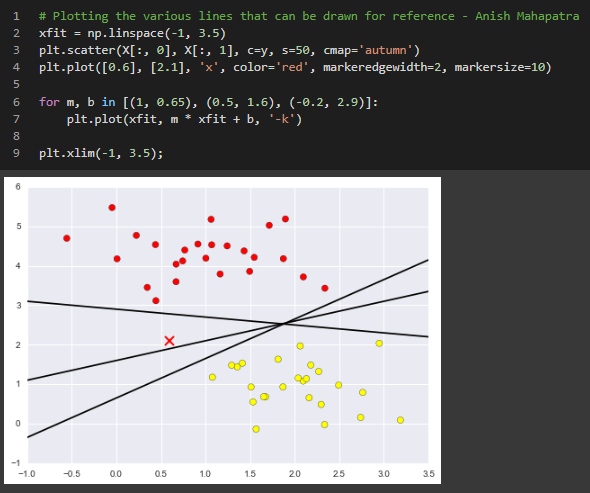
*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

Furthermore, our dataset will look something like this, where we would like to showcase the linear separation of the data.



*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

Now, if we were to use a linear discriminative classifier, we would attempt to find an optimal straight line between the two sets of data such that we are able to segregate the datasets. There are various lines that can be drawn to segregate the datasets.

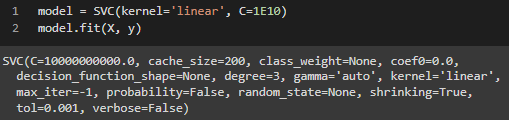


*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

Confused about which one to choose? Remember what we discussed in the previous sections? Our aim is to **maximize the margin**. In the next section, we will discuss exactly this.

## Maximize the margin

Now, we need to fit the linear support vector machine so that we are able to plot the optimal hyperplane to get the best fit model. In this case, we will be using the linear kernel, as the points in the X and Y axis have a linear relationship.

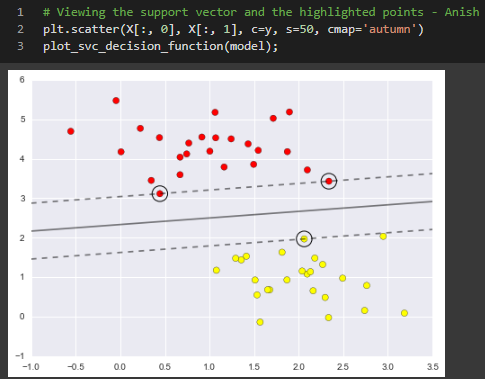


*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

We use the linear kernel within the support vector classifier (svc) from the Sciket learn package to segregate the datasets appropriately. The aim of the dividing decision boundary is to maximize the margin between the different groups of points. Some of the points touch the line and are indicated separately. These points are critical and are known as support vectors, and they are stored in the support\_vectors\_ attribute of the function.

## Fit the support vector machine classifier

Based on hyperparameter tuning, it is to be decided what the best possible model for the given dataset will be. We notice the support vectors here, and the position of the straight dividing line (called hyperplane for n-dimensional data) will change based on how the margins can be maximized.

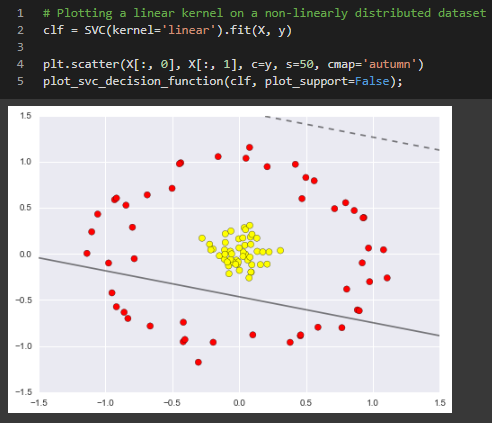


*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

Based on the parameters and the number of rows in the train and test data, the position and accuracy of the svm model will vary.

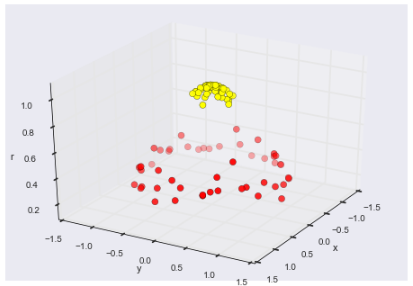
## Decide the kernel type based on the data distribution type

Based on the distribution of the data, it is possible also to have a non-linear dataset distribution that can be solved using other kernels. For instance, if we were to use a linear kernel on a non-linearly distributed dataset, we would see a plot that looks similar to the following.



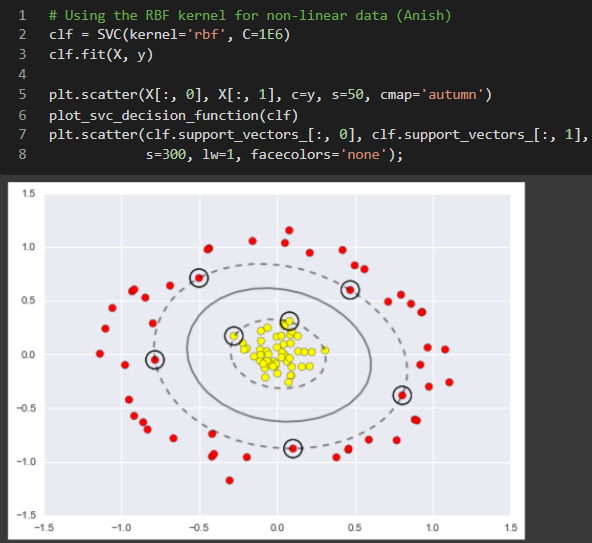
*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

Whereas, if we were to project and transform the two-dimensional data onto a three-dimensional space, it would look like the following.



*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

Here, in this case, if we used the RBF kernel, the plot would look like the below image, where we have successfully segregated and mapped the hyperplane back to the original points.

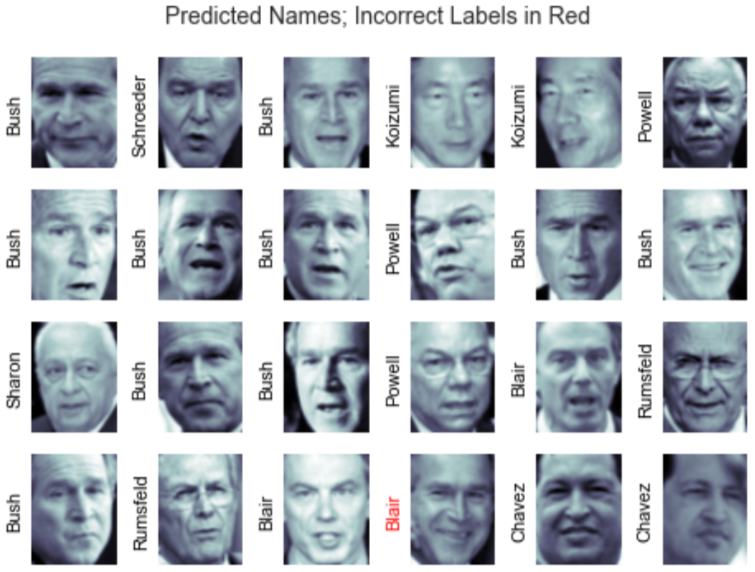


*Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)

In this section, we successfully went over some simple Python code to generate relevant datasets and showcased how support vector machines can be used to generate fairly accurate models with minimal computation using the kernel trick. In the next section, we will go over some of the applications of support vector machines which can also be learnt via the [best online data science courses](https://www.knowledgehut.com/data-science-courses).

# Applications of Support Vector Machine

In this section, we will go over some of the use cases of support vector machines.

* **Email classification**: A support vector machine can be used for email classification, where we decide if an email is a spam or ham
* **Face Detection**: Leveraging svm, we can perform face recognition, where we train the model on the dataset, and we can predict. In the below image, we can see the training dataset.   
    
  *Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)  
  Moreover, in the below image, we can see the test dataset, where the text in red indicates that the image has been incorrectly predicted.   
     
  *Image by* [*Author*](https://www.linkedin.com/in/anishmahapatra/)  
  We are also able to get the metrics such as precision, recall and f1-score for the same.
* **Text categorization:** Categorization of both inductive and transductive models is used for training, and it uses different scores generated to compare with the threshold value.
* **Handwriting recognition**: SVM can also be used for handwriting recognition, where we are able to convert hand-written text to machine-readable text.
* Bioinformatics: This includes cancer classification and protein classification, where we use SVM to identify the classification of patients and genes based on biological markers.

# Advantages and Disadvantages of Support Vector Machine

In this section, we will go over the advantages of SVM and disadvantages of SVM.

## Advantages of SVM

* When there is a clear margin of separation between various classes, SVM works well
* Memory efficiency is one of the key advantages of SVM as it uses a subset of training points as part of the decision function support vectors)
* SVM tends to be an optimized algorithm when the data exists in a high dimensional space
* It works well when there is a higher number of columns than the number of rows
* It is possible to use various kernel functions to make better models

## Disadvantages of SVM

* SVM is able to achieve better results with smaller datasets with higher dimensions, as large datasets may take a longer time
* We need to be aware that when we have a large number of features, we need to avoid overfitting in the data and regularization of terms is critical
* Due to the kernel trick, SVM works by projecting the data into a higher dimension, so there is no probabilistic explanation to perform classification. This can be done, but it uses a computationally intensive five-fold cross-validation.

# Conclusion

In this blog, we went over end-to-end questions that will be asked in an interview. We learn how svm can be used in various use cases on linear and non-linearly separable data. We understood the kernel trick and also understood the various terminology associated with support vector machines. We also went over some simple coding examples and understood how we could work to maximize the margin with the help of support vectors. Lastly, we went over some of the applications of the SVM and also looked at its pros and cons of svm. Great job. Please feel free to post your questions in the comments

# FAQs

## What are support vector machines with examples?

A support vector machine is a set of supervised learning models that can be used for classification as well as regression. It works well on high dimensional data and has fairly high accuracy and minimal computation time, especially with smaller training datasets. Unfortunately, it does not provide a probabilistic estimation of the points. It is used for linearly and non-linearly separable data and is used in cases such as email classification, text categorization, face detection, and handwriting recognition.

## Which type of classifier is SVM?

SVMs are **maximal-margin classifiers**, as compared to some other algorithms, such as Naïve Bayes, which are probabilistic-based classifiers. Besides this, there are other kernel functions we can use as well.

## Which is better, SVM or neural network?

Depending on the use case, the model needs to be chosen. The prediction time for neural networks is faster than SVM. The processing for the parameters of SVM increases linearly with the increase in the size of the input. In many use cases, neural networks perform better but can be computationally expensive.

## What are the advantages of SVM?

There are multiple advantages of SVM. It has better memory efficiency and works well with high-dimensional data (Where the number of columns is more than the number of rows). It also works well when the size of the data is relatively small. It also has the advantage of the kernel trick, where various kernels can be used to segregate the data better.

## Can we use SVM for regression?

Yes, SVM can be used for regression, and the Support Vector Regression (SVR) package in Python can be used for the same. It uses similar principles as support vector machines but for regression problems.

# Author

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