Linear Support Vector Machines Explained

With video explanation | Data Series | Episode 9.1

Mazen Ahmed · Follow

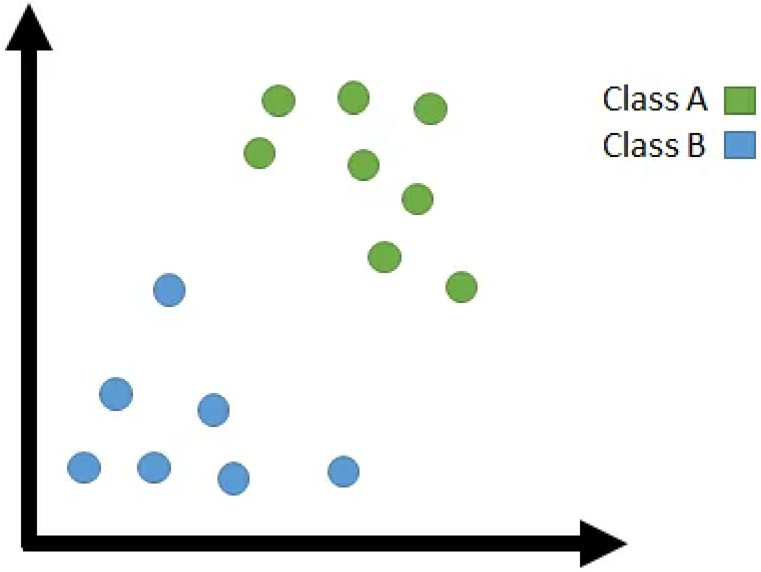
5 min read · Feb 11, 2021

What are Support Vector Machines?

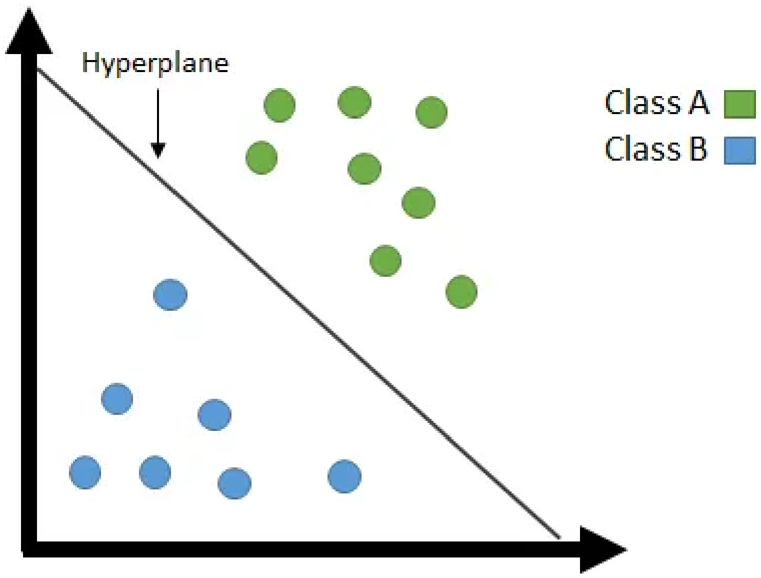
Support Vector machines are a common **supervised** machine learning algorithm used in both **classification** and **regression** problems, however are most commonly used for classification which will be the focus for this article.

Overview

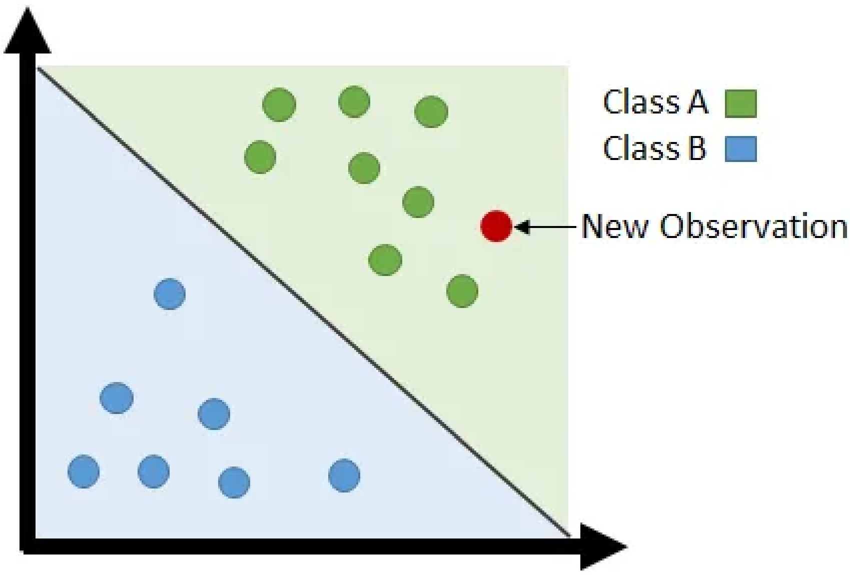
The job of a support vector machine for classification problems is take **labelled data** such as the following:



and determine a **hyperplane** that separates the data:

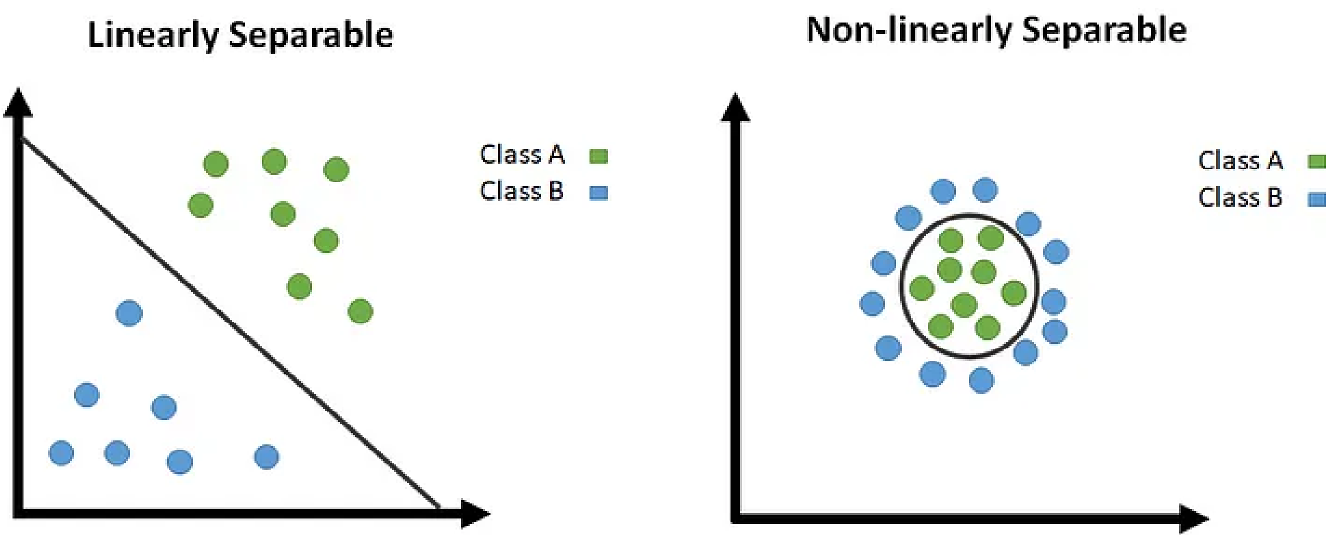


We can then use this hyperplane to make **predictions** for which class a **new data point** belong to:



Since our new observation lies **above the hyperplane** we predict this observation belongs to **Class A.**

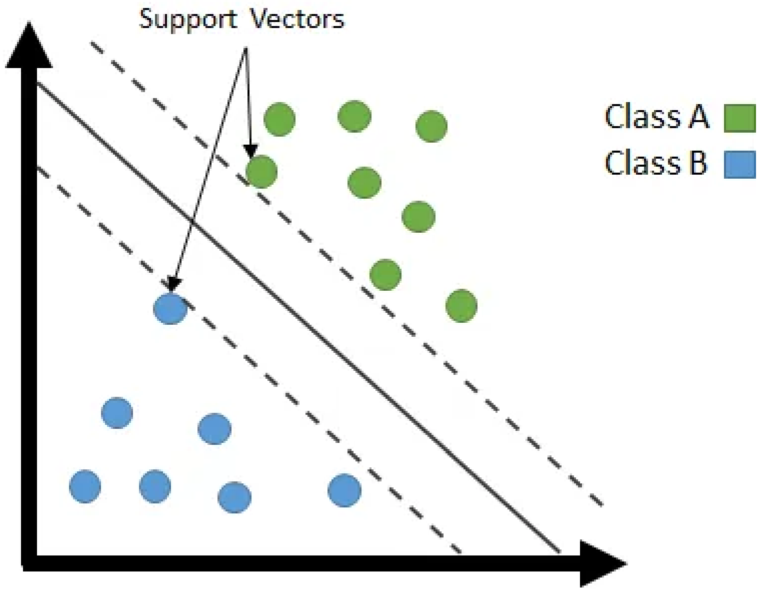
Here, we will be focussing on **linearly separable** support vector machines. The difference between linearly separable and non-linearly separable data is illustrated below:



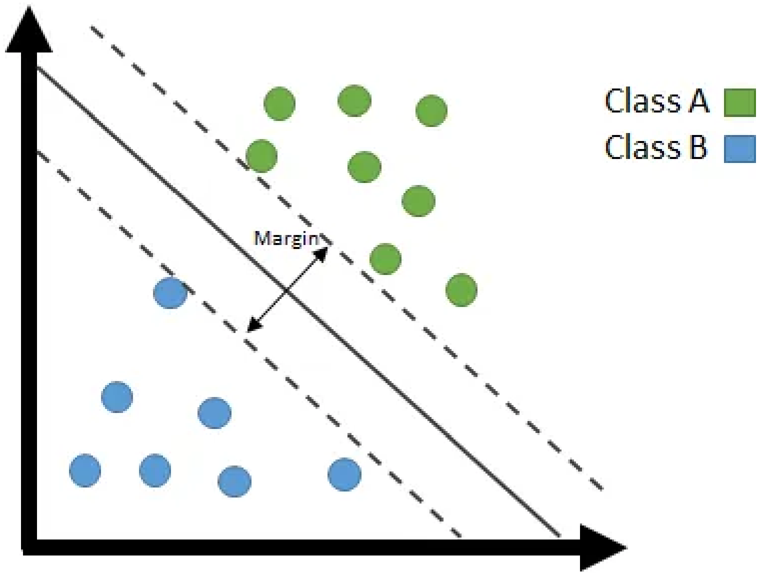
Terminology

Before we go on to see the general method of calculating the hyperplane we must first be aware of some support vector machine terminology.

Support Vectors: Vectors or points that are closest to our hyperplane from each class, if these vectors were to be removed the position of our hyperplane may change.



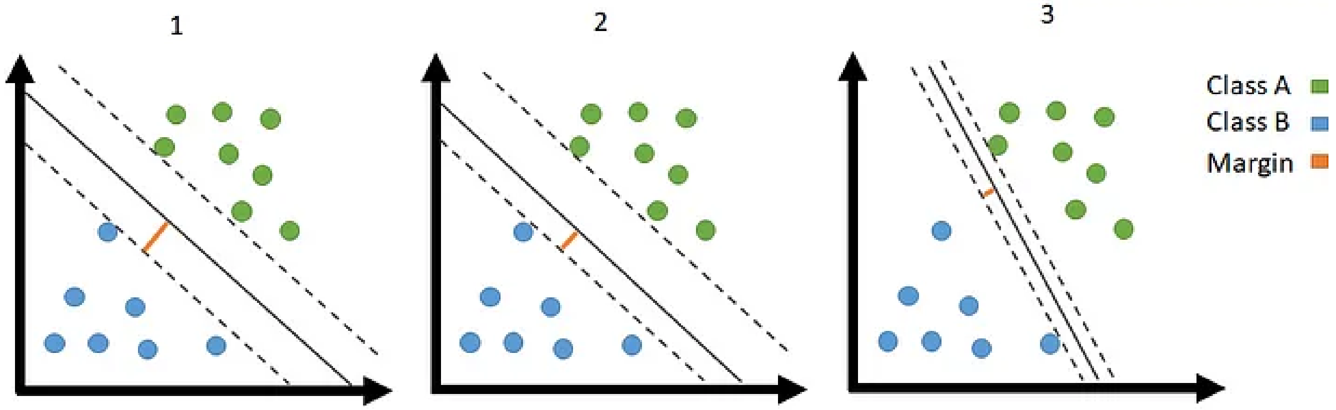
Margin: The distance our hyperplane is away to the **closest** vector. In this case the hyperplane is at the mid-point of the two support vectors so the margin values are the same.



Hyperplane (SVM): A decision boundary used to separate and classify data. In 2-dimensional space (2 features) a hyperplane is 1 dimensional which is a line (the case we have above). In 3-dimensional space (3 features) a hyperplane is 2 dimensions, a plane. In general a Hyperplane is n-1 dimensions where n is our number of features.

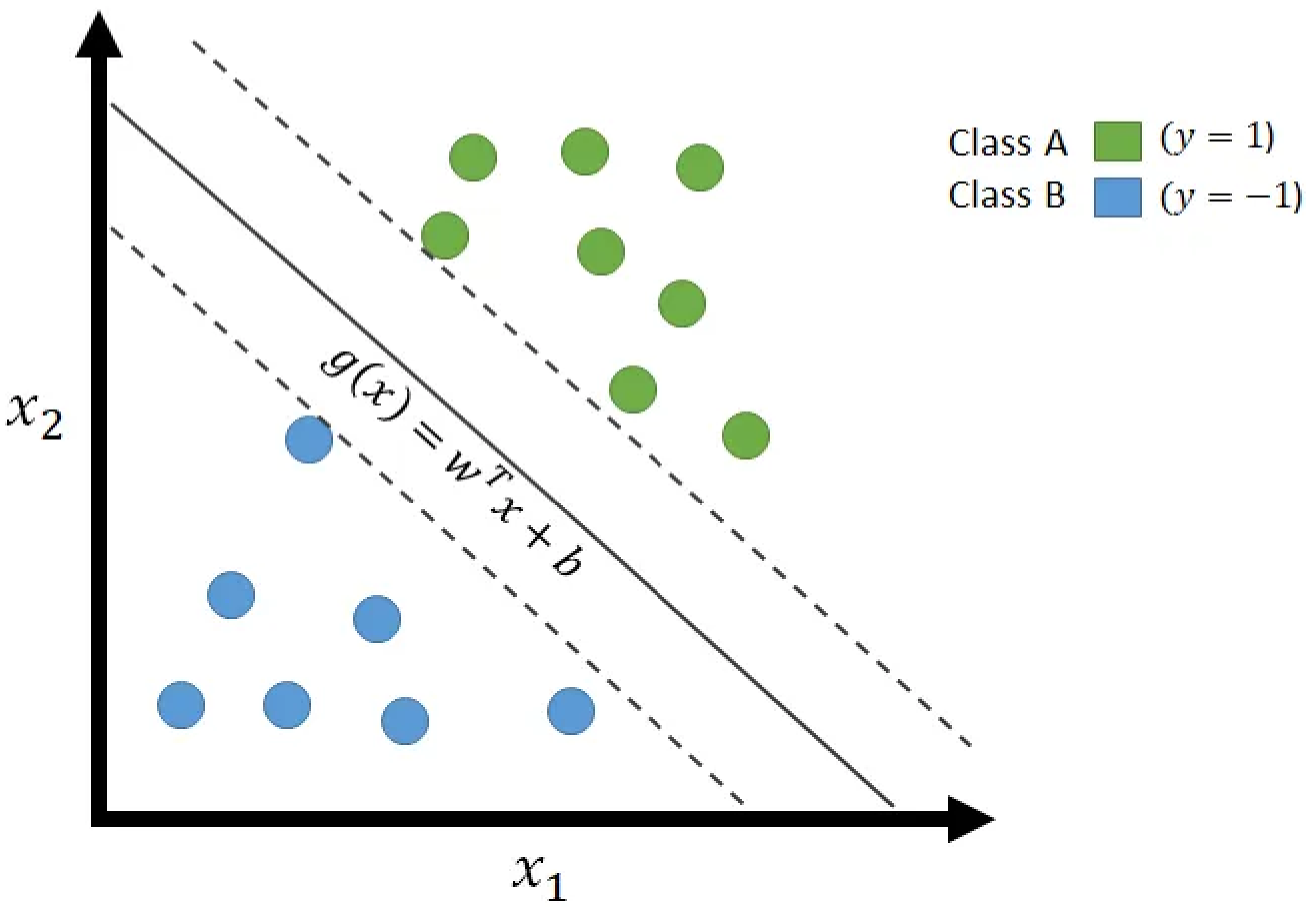
Calculating the Hyperplane

Take a look at the following three hyperplane separating our data:



Our optimal hyperplane is given by the hyperplane with the **largest margin** which in this case is shown in **graph 1** above.

Our **Hyperplane** takes the following formula:



$$g(x) = w^T x + b = w_1 x_1 + w_2 x_2 + b$$

where

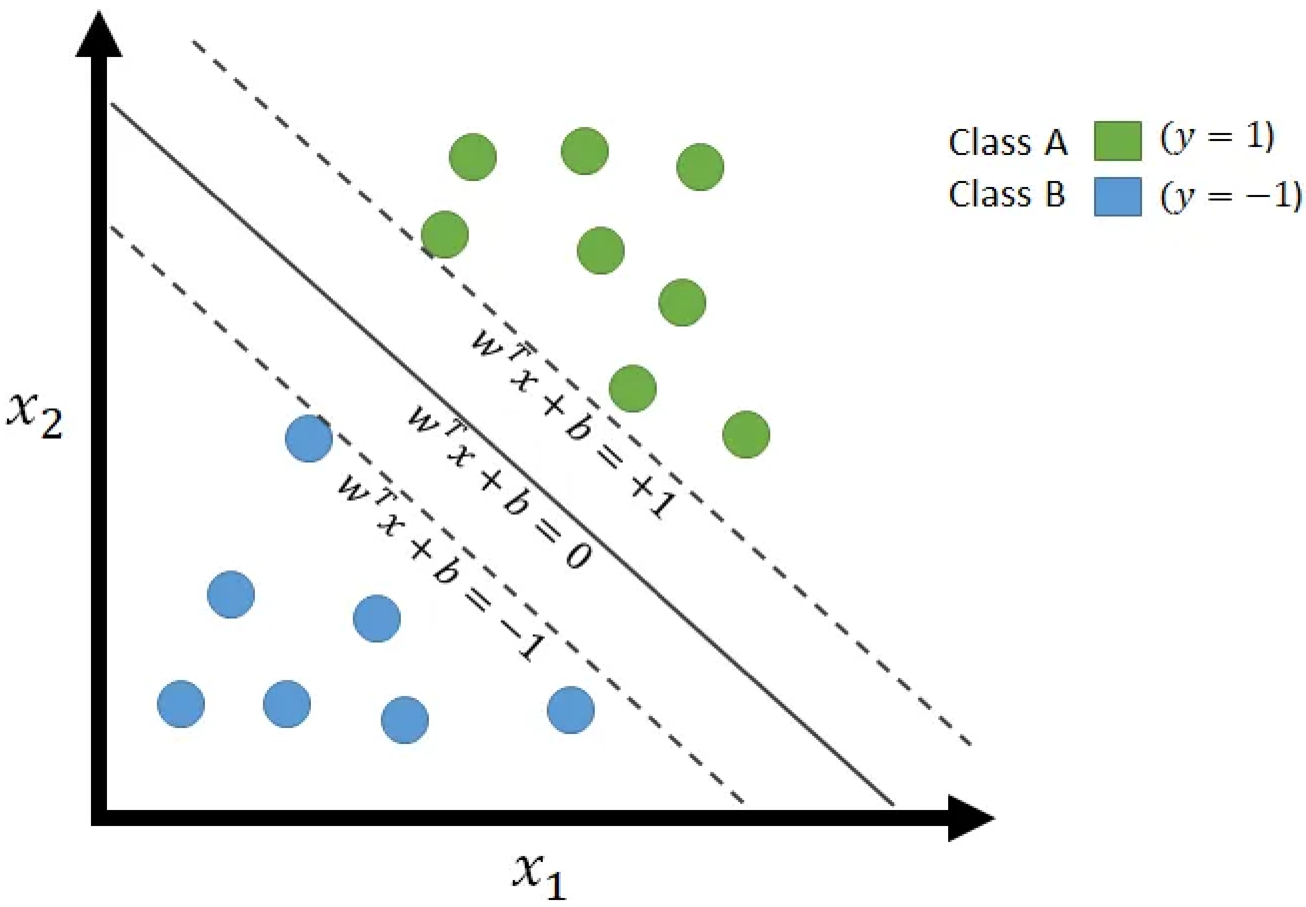
- w is a vector of weights of parameters

- **b** is our bias term or intercept
- x is our input vector, in the above case we just have two features x_1 and x_2 but we may have hundreds or more.

All data points x that belong in class A have label y = 1.
All data points x that belong in class B have label y =-1.

We give an objective that our hyperplane:

- predicts a value **greater than or equal to 1** for all data points in **Class A (when y = 1).**
- Predict a value **less than or equal to -1** for all data point in **Class B. (when y = -1).**



Mathematically:

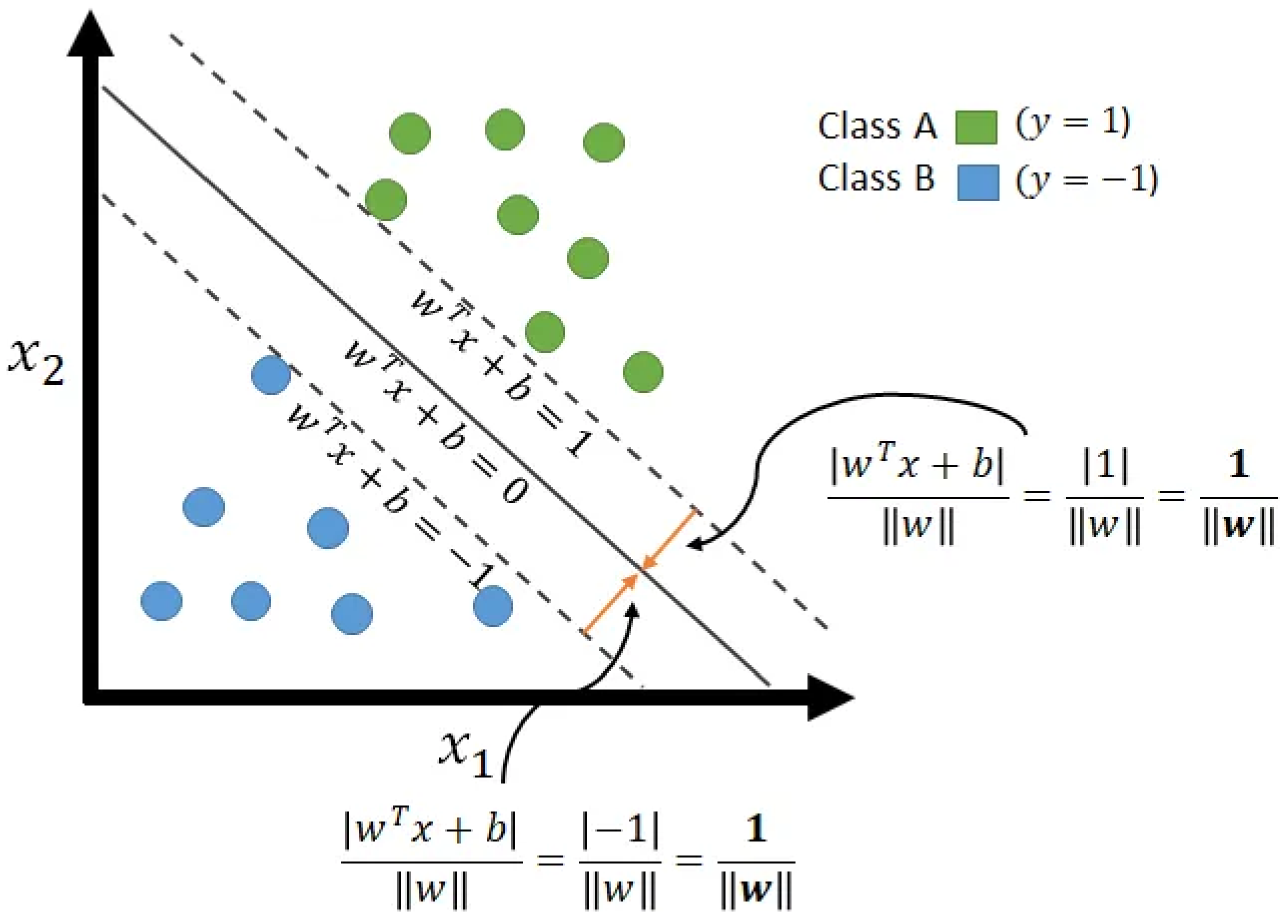
$$w^T x_i + \mathrm{b} \geq 1 \quad \text{for all } (y_i = 1)$$
$$w^T x_i + \mathrm{b} \leq -1 \quad \text{for all } (y_i = -1)$$

We can combine the above conditions into one to give:

$$(w^T x_i + b) y_i \geq 1$$

Keeping this condition in mind, we will now calculate algebraically our margin.

Using the formula for the distance between a **line** and a **point**:



giving a total margin of:

$$\frac{1}{\|w\|} + \frac{1}{\|w\|} = \frac{2}{\|w\|}$$

Remember that best hyperplane is that with the **largest margin**. So to maximise the margin above we must **minimise**:

$$||w||$$

We take the square of this value and half it, to ease solving this minimisation problem, and are left with the following minimization problem with the condition identified before:

$$\min \frac{1}{2}||w||^2$$

$$\text{Such that: } (w^T x_i + b)y_i \geq 1$$

$$\text{for } \ \mathrm{i} = 1, \dots, \mathrm{m}$$

This is a non-linear **programming problem**, which can be solved using Karush-Kuhn-Tucker conditions. Applying the aforementioned method, we obtain the conditions:

$$\mathrm{w} = \sum_{i=1}^{m} \lambda_i y_i x_i \qquad \text{and} \qquad \sum_{i=1}^{m} \lambda_i y_i = 0$$

where **m** is our number of training examples and λ is our Lagrangian multiplier.

Read more about how these conditions are obtained here:
https://www.csd.uwo.ca/~xling/cs860/papers/SVM_Explained.pdf

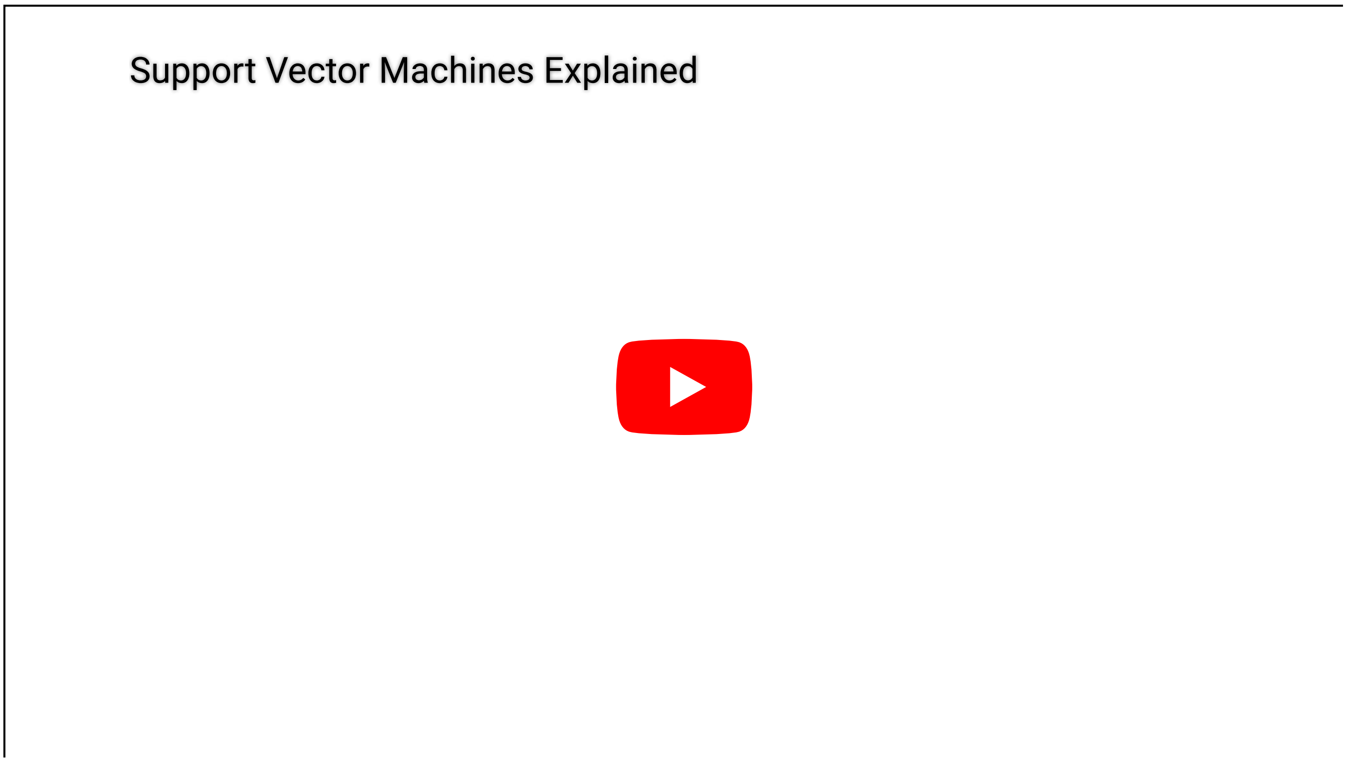
We use the conditions above to obtain optimal values for **w** and **b** given training data *x* and labels **y**, which gives us the equation of the hyperplane with the **largest margin** as required.

Summary

- Support Vector Machines are commonly used in classification problems.
- SVMs work by calculating a hyperplane which separates our data with the largest margin. This is done using and optimization technique called **non-linear programming.**
- We then use this hyperplane as a decision boundary to classify data into groups.

Prev Episode | Next Episode

If you have any questions please leave them below!



Data Science Machine Learning Support Vector Machine Mathematics AI

Follow

Written by Mazen Ahmed

205 Followers

PhD Student in Bioinformatics that has produced an on-going series of data science tutorials with Python examples.

More from Mazen Ahmed

Mazen Ahmed

Ways to Evaluate your Regression Model

The different methods | Data Series | Episode 10.1

· 5 min read · Oct 6, 2021

9

Mazen Ahmed

Evaluating your Regression Model in Python

Step-by-step follow along | Data Series | Episode 10.2

· 6 min read · Oct 21, 2021

4

Mazen Ahmed

Data Science Project | Clustering Mixed Data

Start to Finish Clustering Analysis | Data Series | Project 3

· 8 min read · Dec 29, 2021

Mazen Ahmed

Logistic Regression in Python

Step-by-step follow along | Data Series | Episode 7.2

· 5 min read · Oct 17, 2020

See all from Mazen Ahmed

Recommended from Medium

Aziz Budiman

Machine Learning in 4 Minutes: Support Vector Machines

In part 2 on this mini-series of #4MinutesML, we will uncover the mysteries behind the use of Support Vector Machines (SVM) and its...

4 min read · Jun 3

19

Zeki

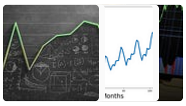
Practical Example of Clustering and Radial Basis Functions (RBF)

Clustering is a technique used in machine learning and data analysis to group similar data points together. The goal of clustering is to...

9 min read · Feb 17

4

Lists

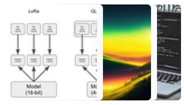


Predictive Modeling w/ Python

18 stories · 57 saves

Practical Guides to Machine Learning

10 stories · 71 saves

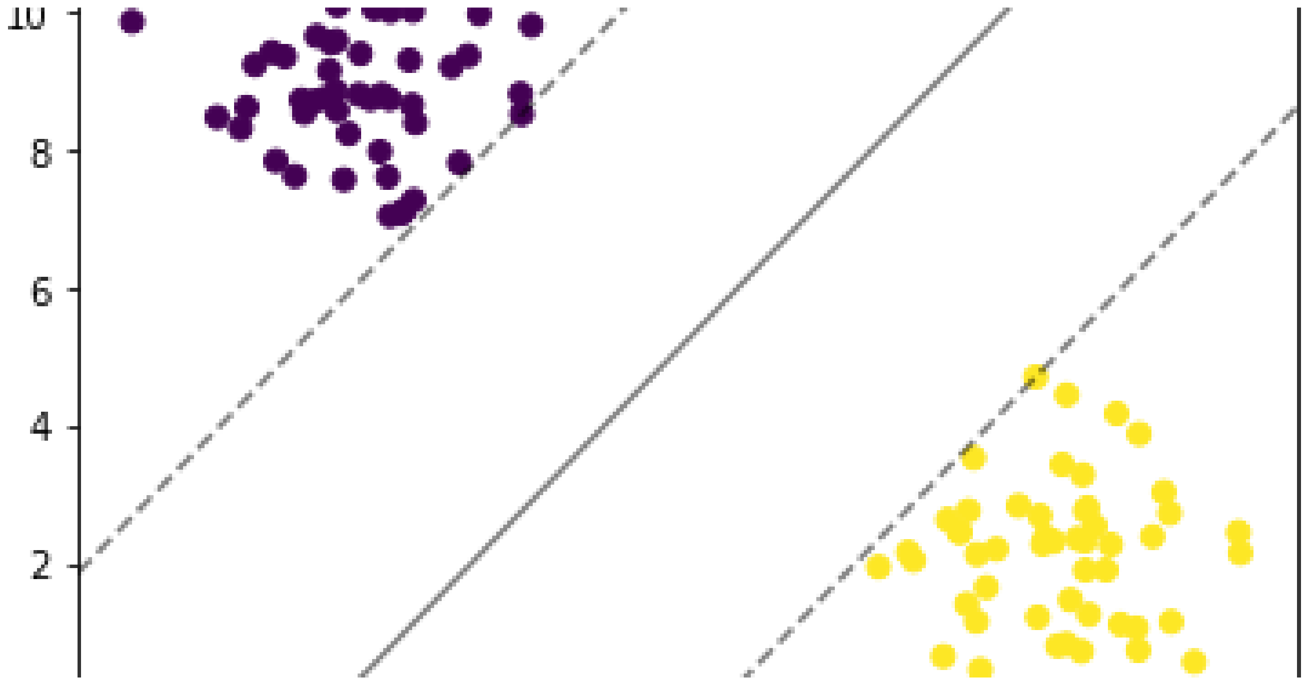


Natural Language Processing

369 stories · 21 saves

The New Chatbots: ChatGPT, Bard, and Beyond

13 stories · 29 saves



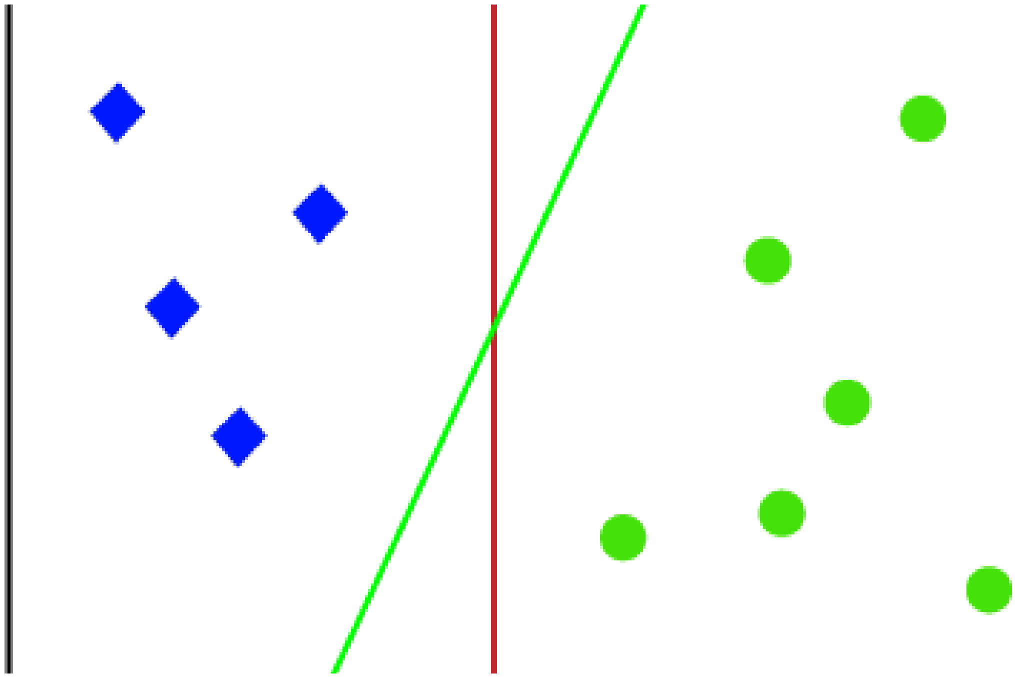
 Sharmasaravanan

Introduction to Support Vector Machine (SVM) in Machine Learning

Support Vector Machine (SVM) is a powerful supervised machine learning algorithm used for classification and regression analysis. SVM works...

3 min read · Mar 11

11

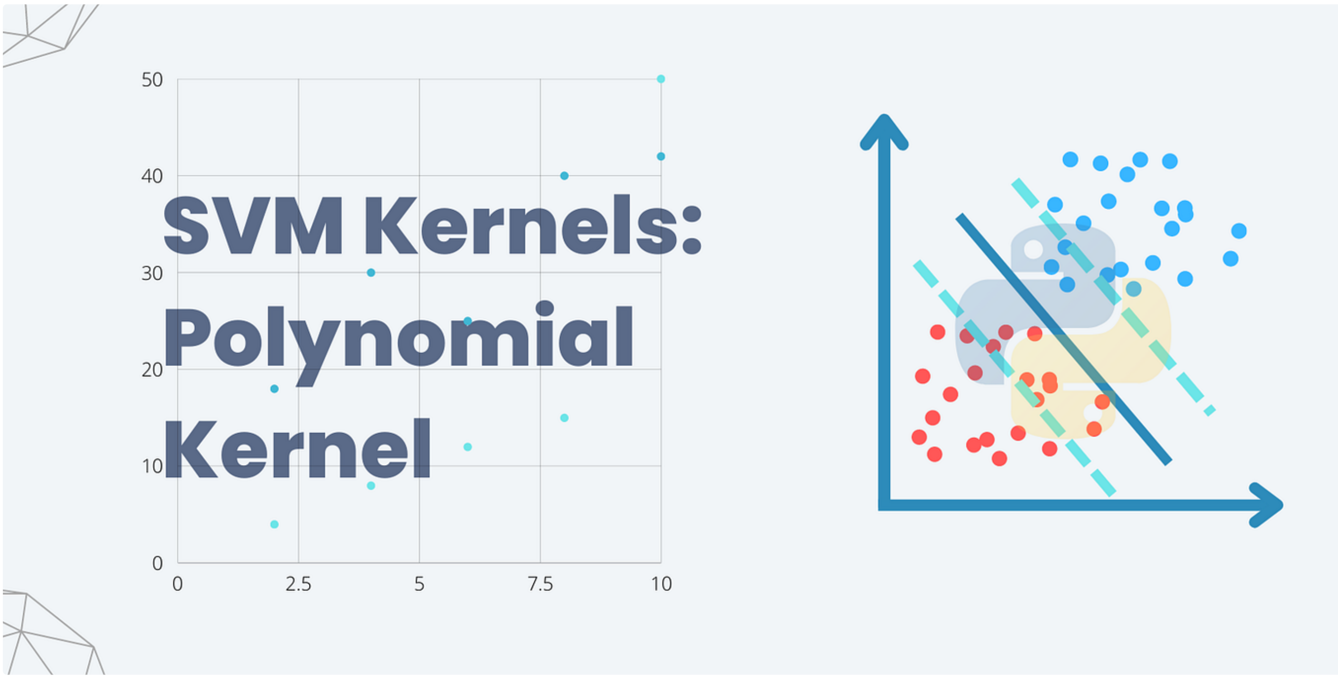


Detheharshada

Support vector machines(SVM)

Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. It is...

4 min read · May 3



Mert Demir in Dev Genius

Machine Learning Algorithm Series Polynomial Kernel SVM: Understanding the Basics and Applications

Understanding the Basics and Applications

6 min read · Apr 7

26

Mohammad Derakhshan

How to implement the shimmer effect using jetpack compose?

In almost every project, we need to load some data and present them to the user. Sometimes, the user must wait a few moments for data to be...

1 min read · Jan 20

151

See more recommendations