# SVM Tutorial

# SVM - Understanding the math - Part 1 - The margin

## Introduction

This is the first article from a series of articles I will be writing about the math behind SVM. There is a lot to talk about and a lot of mathematical backgrounds is often necessary. However, I will try to keep a slow pace and to give in-depth explanations, so that everything is crystal clear, even for beginners.

If you are new and wish to know a little bit more about SVMs before diving into the math, you can read the article: an overview of Support Vector Machine.

## What is the goal of the Support Vector Machine (SVM)?

> *The goal of a support vector machine is to find the optimal separating hyperplane which maximizes the margin of the training data.*

The first thing we can see from this definition, is that a SVM needs training data. Which means it is a supervised learning algorithm.

It is also important to know that SVM is a classification algorithm. Which means we will use it to predict if something belongs to a particular class.

For instance, we can have the training data below:

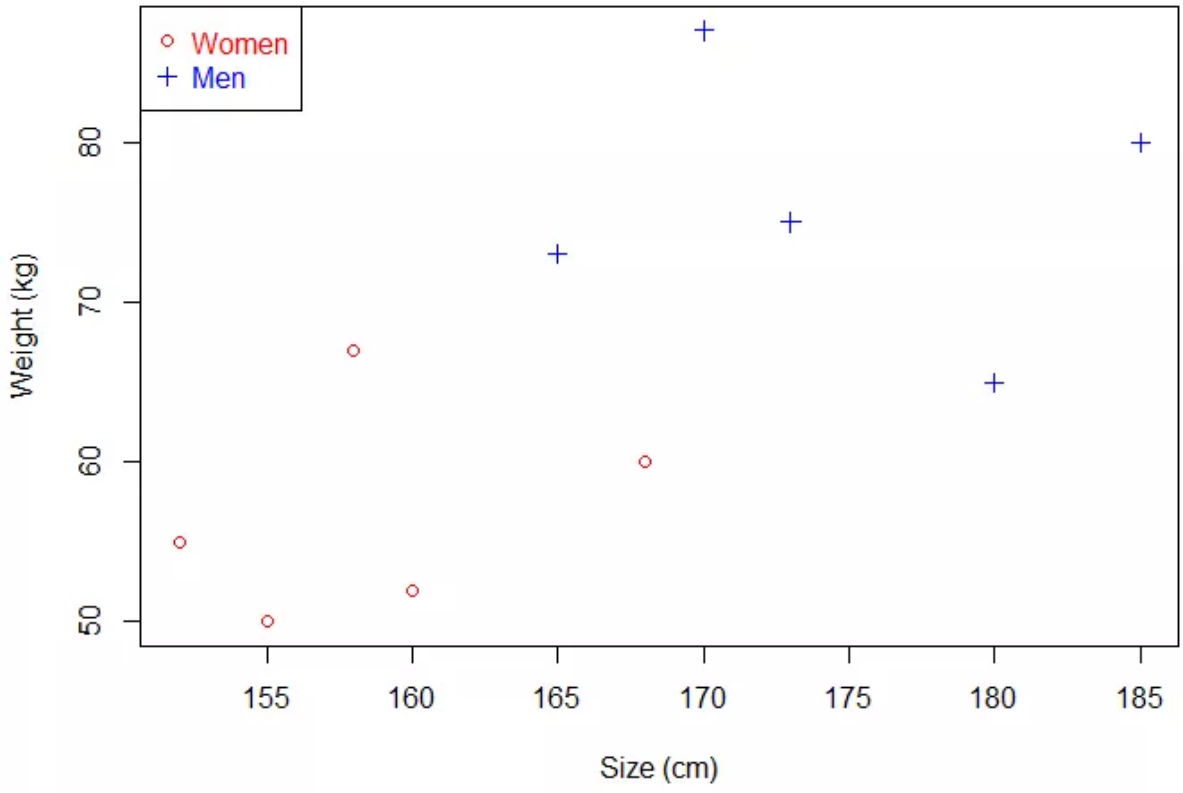



*Figure 1*

We have plotted the size and weight of several people, and there is also a way to distinguish between men and women.

With such data, using a SVM will allow us to answer the following question:

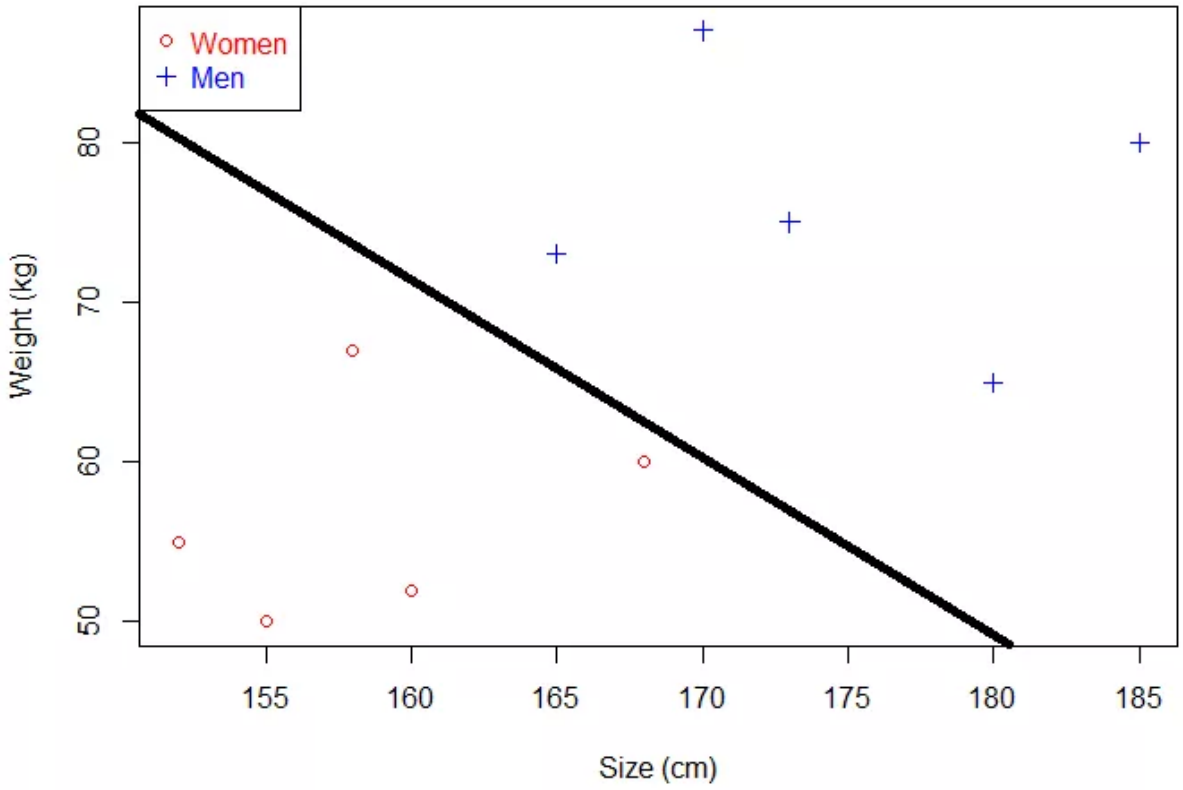
> *Given a particular data point (weight and size), is the person a man or a woman ?*

For instance:  if someone measures 175 cm and weights 80 kg, is it a man of a woman?

## What is a separating hyperplane?

Just by looking at the plot, we can see that it is possible to separate the data.  For instance, we could trace a line and then all the data points representing men will be above the line, and all the data points representing women will be below the line.

Such a line is called a **separating hyperplane** and is depicted below:

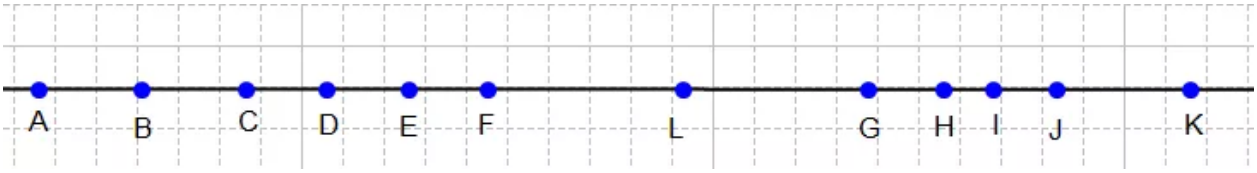



## If it is just a line, why do we call it an hyperplane ?

Even though we use a very simple example with data points laying in $R^2$ the support vector machine can work with any number of dimensions !

**An hyperplane is a generalization of a plane**.

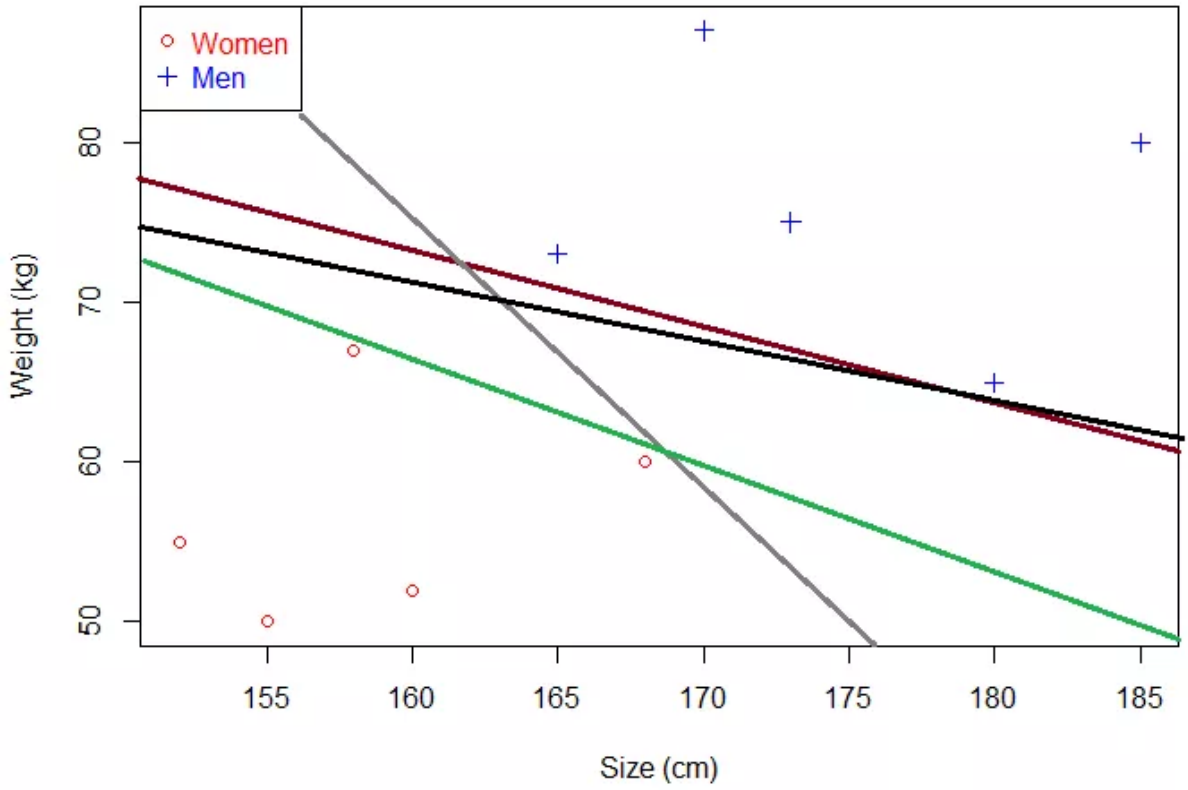
- in one dimension, an hyperplane is called a point
- in two dimensions, it is a line
- in three dimensions, it is a plane
- in more dimensions you can call it an hyperplane




*The point L is a separating hyperplane in one dimension*

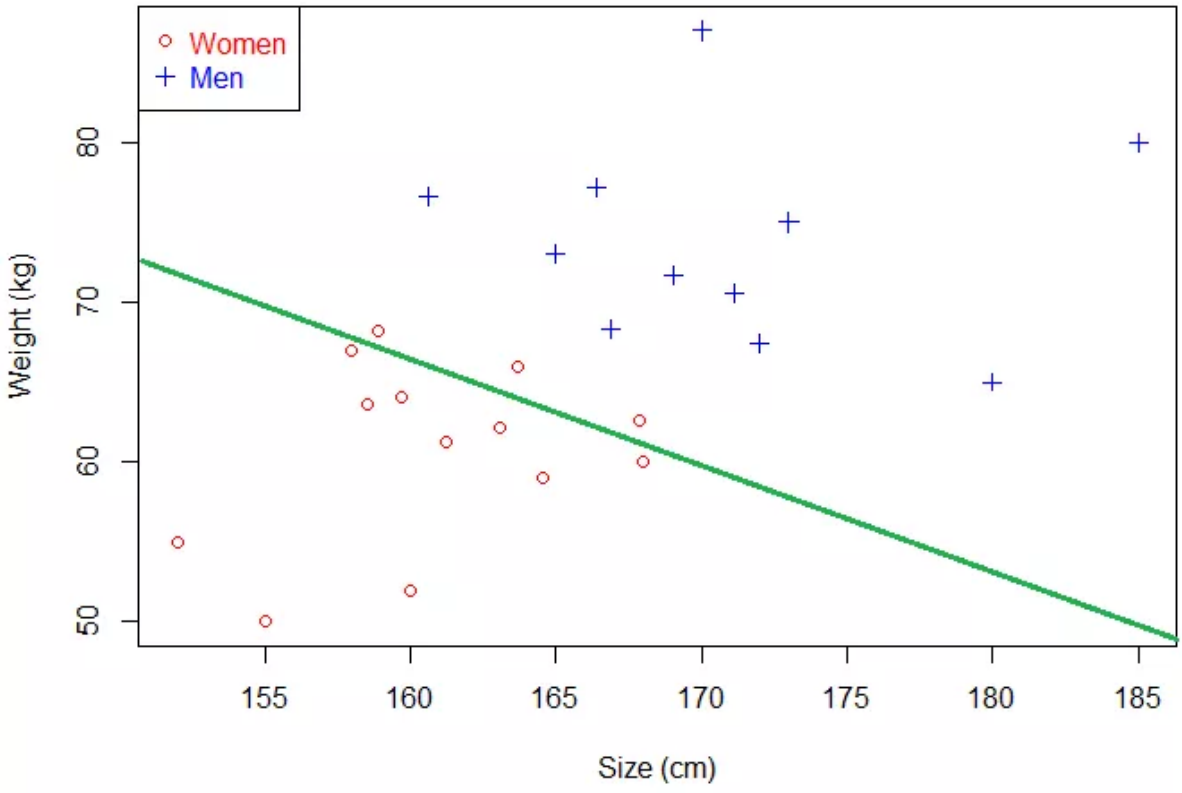
## What is the *optimal* separating hyperplane?

The fact that you can find a **separating hyperplane**, does not mean it is the best one ! In the example below there is several separating hyperplanes. Each of them is valid as it successfully separates our data set with men on one side and women on the other side.




*There can be a lot of separating hyperplanes*

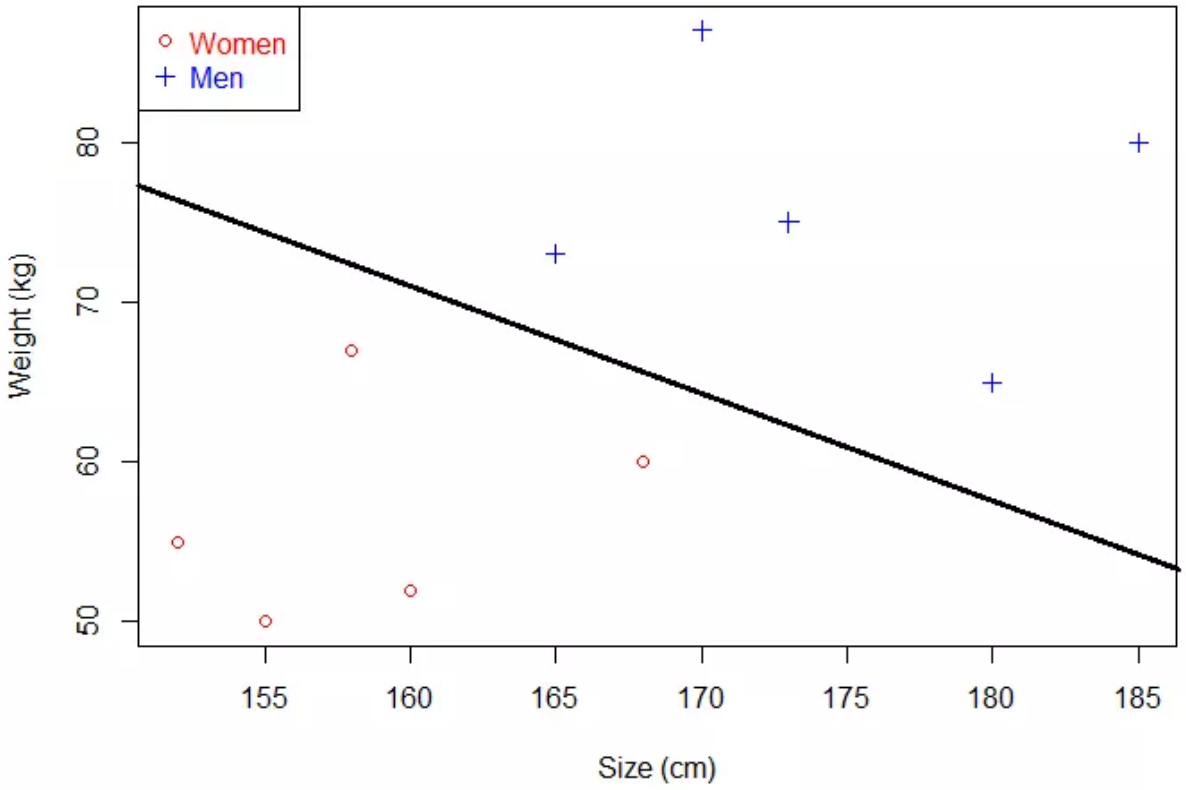
Suppose we select the green hyperplane and use it to classify on real life data.



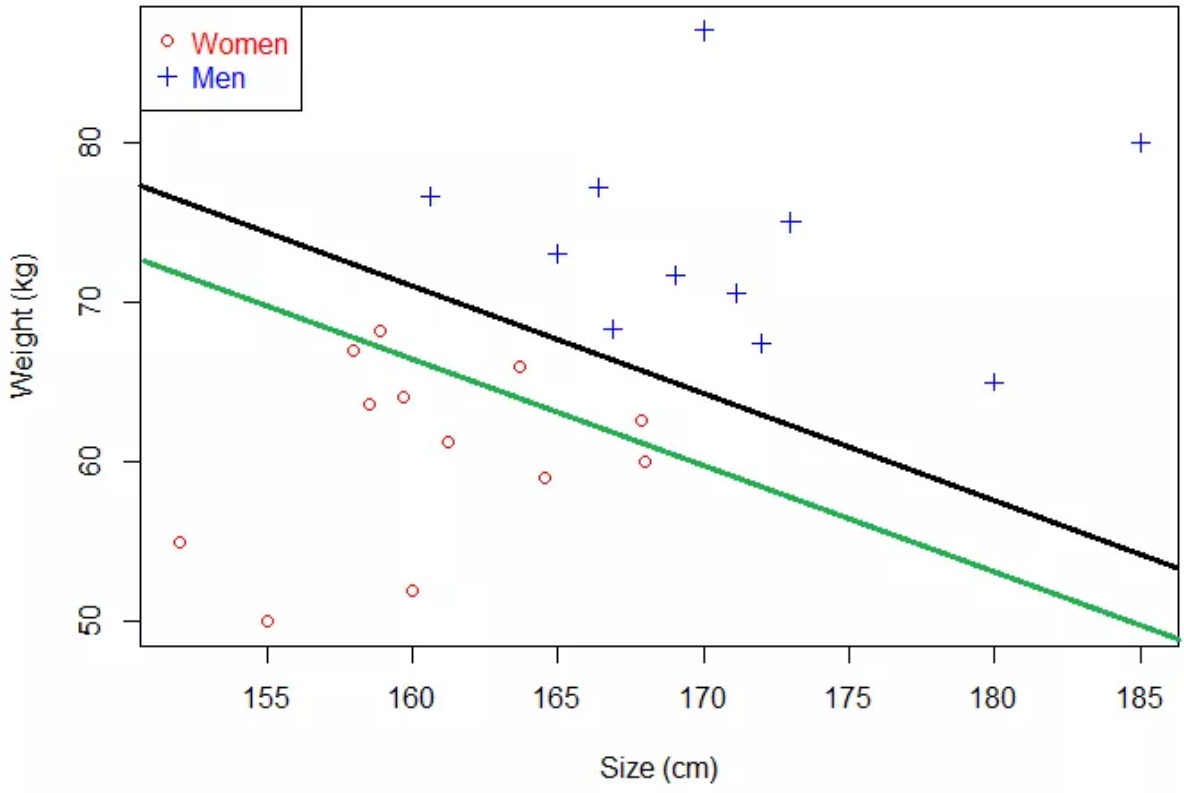

*This hyperplane does not generalize well*

This time, it makes some mistakes as it wrongly classify three women. Intuitively, we can see that *if we select an hyperplane which is close to the data points of one class, then it might not generalize well.*

So we will try to select an hyperplane **as far as possible from data points from each category:**




This one looks better. When we use it with real life data, we can see it still make perfect classification.

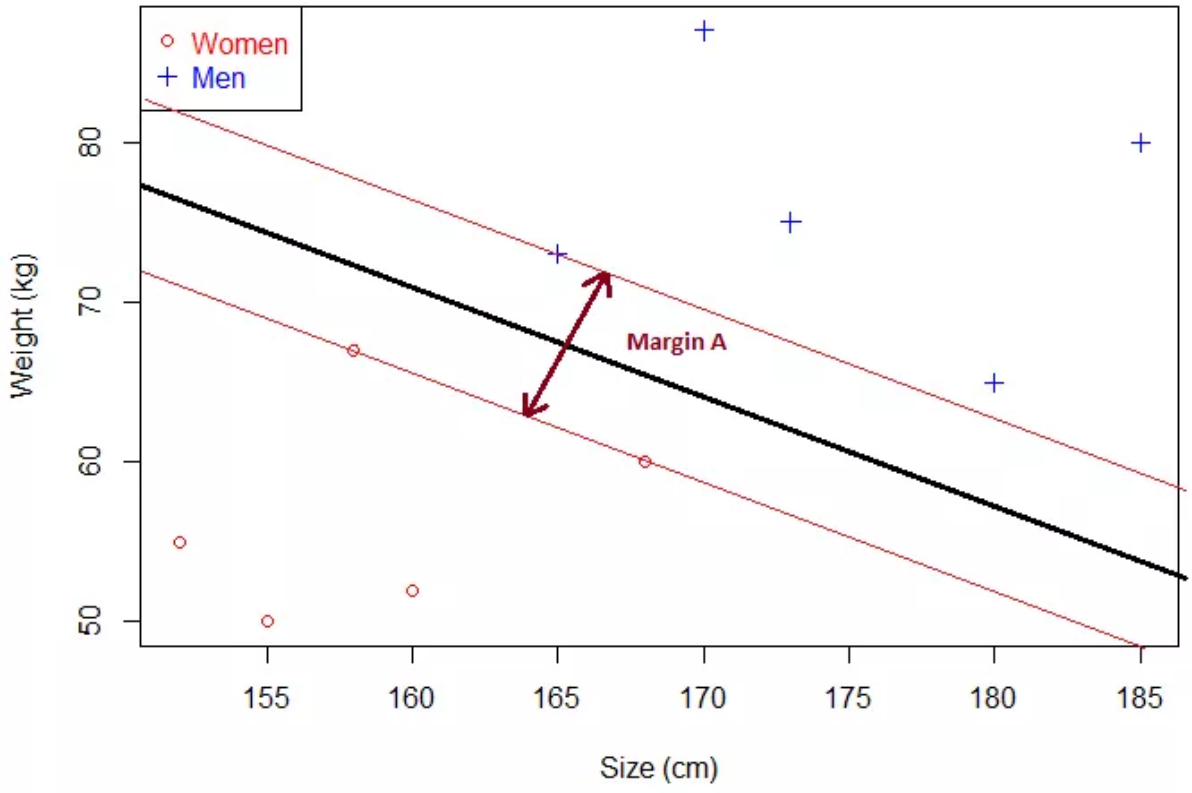



*The black hyperplane classifies more accurately than the green one*

That's why the objective of a SVM is to **find the optimal separating hyperplane**:

- because it correctly classifies the training data
- and because it is the one which will generalize better with unseen data

## What is the margin and how does it help choosing the optimal hyperplane?

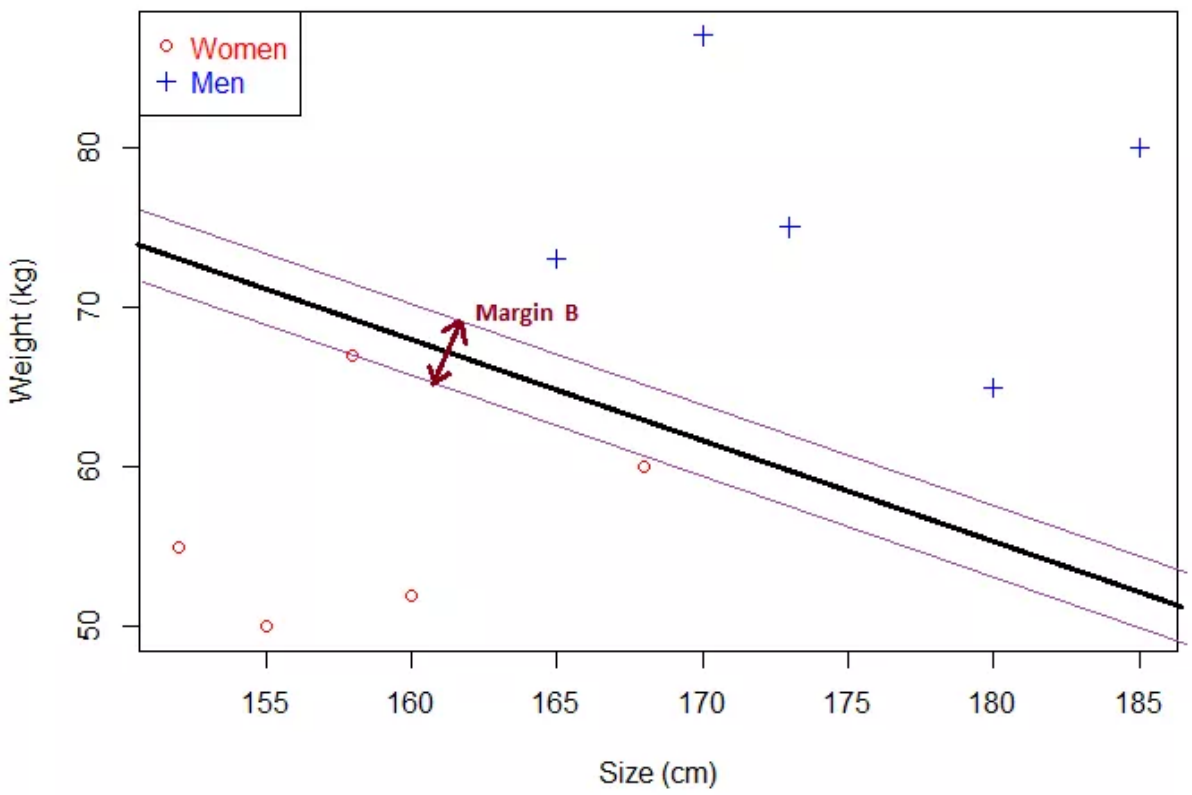



*The margin of our optimal hyperplane*

Given a particular hyperplane, we can compute the distance between the hyperplane and the closest data point. Once we have this value, if we double it we will get what is called the **margin**.

**Basically the margin is a no man's land. There will never be any data point inside the margin.** (Note: this can cause some problems when data is noisy, and this is why soft margin classifier will be introduced later)

For another hyperplane, the margin will look like this :




As you can see, Margin B is smaller than Margin A.

We can make the following observations:

- If an hyperplane is very close to a data point, its margin will be small.
- The further an hyperplane is from a data point, the larger its margin will be.

This means that **the optimal hyperplane will be the one with the biggest margin.**

That is why the objective of the SVM **is to find the optimal separating hyperplane which maximizes the margin of the training data.**

This concludes this introductory post about the math behind SVM. There was not a lot of formula, but in the next article we will put on some numbers and try to get the mathematical view of this using geometry and vectors.

If you want to learn more read it now :
SVM - Understanding the math - Part 2 : Calculate the margin

Alexandre KOWALCZYK

*I am passionate about machine learning and Support Vector Machine. I like to explain things simply to share my knowledge with people from around the world. If you wish you can add me to linkedin, I like to connect with my readers.*

This entry was posted in Mathematics, SVM Tutorial and tagged hyperplane, math on November 2, 2014 [https://www.svm-tutorial.com/2014/11/svm-understanding-math-part-1/] .

## 101 thoughts on “SVM - Understanding the math - Part 1 - The margin”

David Harper CFA FRM (@bionicturtle)
November 6, 2014 at 2:50 am

Great intro, who are you? Thanks!

**Alexandre KOWALCZYK** Post author
November 7, 2014 at 11:09 pm

Thanks. I just found out my name was not visible on the site. I am passionate about machine learning and I work as a software engineer in a financial firm. 🙂 I added an author box. You can check my profile on various social media now.

**Chetan Yadati**
May 11, 2016 at 8:18 am

Hi Alexandre,
Thanks for writing up this piece( and the following ones) on SVM. I really like the fact that you have explained all

the essentials along with the SVM. It is comprehensive and complete. I could not have wished for a better write-up on the subject.

**Peristic**
October 7, 2016 at 6:49 pm

Hi Alexandre thank you so much you're awesome!!

**Renny Varghese**
December 3, 2016 at 7:02 pm

Could you explain how SVM works for multiple classes? How would it work for 9 classes? I used a function called multisvm here: http://www.mathworks.com/matlabcentral/fileexchange/39352-multi-class-svm but I'm not sure how it's working behind the scenes. Everything I've read online is rather confusing.

**Alexandre KOWALCZYK** Post author
December 4, 2016 at 6:46 pm

Hello Renny, I will dedicate a full chapter of my upcoming (free) ebook to explain multi-class SVM. If you wish you can register at the end of this series so you will receive an email when it is available. 😉

**Anmol Biswas**
December 22, 2016 at 5:28 pm

For N-class classification, what are the other possibilities excluding N one class vs the rest classifiers ? In terms of SVMs

**Alexandre KOWALCZYK** Post author

December 23, 2016 at 1:17 pm

One-vs-One, Crammer & Singer, DAGSVM are alternative approaches for multi-class SVM.

---

Jose
November 7, 2014 at 10:58 pm

This is one of the best tutorials out there. Very well explained. Are you thinking on including some illustrative code on R?

Thanks!

---

**Alexandre KOWALCZYK** Post author
November 7, 2014 at 11:13 pm

Thanks you very much Jose. Right now I am concentrating on the math, but I might add some R code along the way to illustrate some part. I intend to write SVM tutorial like the one about text classification but in R this time as it is widely used.

---

**Arun**
July 3, 2015 at 12:13 pm

Great article. Thanks for such a nice article. I have a fundamental question. The classification here are men and women and hyperplane separates them. If my classes are age say between 1-20 , 21-50 , 51 -75 can we still use svm and how will we model hyperplane.

---

**Alexandre KOWALCZYK** Post author
July 3, 2015 at 11:17 pm

In your example you have what we call a multiclass classification problem. For each person you can assign one of the three classes. The common approach using SVM is the one-vs-rest. So if you have three classes A, B and C you will train three models A vs (B and C), B vs (A and C) and C vs (A and B). Here is an example in python using a linear SVM.

---

Nishant Aggarwal
July 10, 2015 at 12:44 pm

AWESOME ! AWESOME ! AWESOME ! AWESOME !

Thanks,
Nishant Aggarwal

---

**Alexandre KOWALCZYK** Post author
July 11, 2015 at 3:11 pm

Thanks 😉

---

Patrick
July 10, 2015 at 5:30 pm

Hi Alexandre, I would like to use SVR in my thesis to predict a response with 3 causal factors. What is your opinion, does R do it as u showed in your example. Iam pretty new in machine learning

---

**Alexandre KOWALCZYK** Post author
July 11, 2015 at 3:10 pm

Hello Patrick. This is a pretty broad question. R is as good as another language for machine learning. Most often people use R, Python or MatLab. Anyway you can basically do SVR with any language. I don't really understand what you mean by "predicting a response with 3 causal factors". The first thing you have to consider is whether

you need to do a classification or a regression. Then you can pick the appropriate algorithm (SVM for classification or SVR for regression).

---

Patrick
July 13, 2015 at 12:45 pm

Thank you for the response, causal factors I mean I have 3 independent (causal) variables and one dependent variable. So I wd like to use SVR

---

Ahila
September 12, 2015 at 1:53 pm

Hi,
I am new to machine learning and classification.
Do you mind giving an example of EEG classification tutorial

---

**Alexandre KOWALCZYK** Post author
September 14, 2015 at 8:07 pm

I never done EEG classification before. Maybe this video will help you. 😉

---

Emmanuel
November 6, 2015 at 3:29 am

I have 6 classes to classify from and Orthophoto. and Have generated the HSV from the Image including the Mean RGB for each color code.. give these information how to I find the best Optimal Hyperplane using R. Advice me.. Thank you

---

**Alexandre KOWALCZYK** Post author
November 8, 2015 at 4:37 pm

Hello. I am afraid your question is too broad. You have to try for yourself. If you have problems programming, try asking on stackoverflow.

---

Sri Suwarno
November 11, 2015 at 6:53 am

Dear Alexandre, thank you for your clear tutorial. It helps me understanding SVM.

---

Rahul
November 13, 2015 at 10:12 pm

Thank you Alexandre. This is very helpful. SVM was never so clear to me before. However, this clarity on SVM brings me to another question. How is SVM different from the Discriminant analysis. I understand that discriminant analysis as well tries to find a discriminating line which maximizes the distance between points which belongs to different categories. Does the difference lies in the way algorithms are implemented or there is something more to it? Are the application areas of SVM and Discriminant analysis different?

---

**Alexandre KOWALCZYK** Post author
November 15, 2015 at 9:15 pm

Hi Rahul. Thank you for your comment. I don't know a lot about LDA but I found this quora answer which might help you to understand the difference between SVM and LDA

---

**Rahul**
November 19, 2015 at 8:00 am

Thank you Alexandre.

Prihananto Joko
November 27, 2015 at 6:52 pm

hey brother.. thank you ,,, you safe me... You make everything about SVM clear

Akshay
January 14, 2016 at 12:18 pm

Very nicely explained, thank you...

Gurpreet
January 21, 2016 at 8:25 am

Thanks for clarity.....

linmaung
January 23, 2016 at 6:32 pm

thank you for your SVM tutorial.Your explanation is excellence to understand for me.

Abhilasha
February 1, 2016 at 4:45 am

Hi, I love your explanation. I have a silly doubt. What exactly is R^2 that you've mentioned here.
"With data points lying in R^2..."

**Alexandre KOWALCZYK** Post author
February 1, 2016 at 11:28 pm

R^2 represent the euclidian plane. It comes from set theory. If R is the set of all real numbers, then R^2 = R X R is the cartesian product of R. That is, R × R is the set of all ordered pairs whose first coordinate is an element of R and whose second coordinate is an element of R.

Chungkwon Ryu
February 3, 2016 at 7:31 am

Hello! This tutorial was very good for me. So I want to introduce this tutorial to my co-workers by using some slides. Is it possible?

**Alexandre KOWALCZYK** Post author
February 3, 2016 at 8:26 pm

Hello. No problem for me as long as you indicate it came from this site. 😉

Manny Grewal
February 4, 2016 at 9:27 am

Thank you. this is the best tutorial of a newbie. I saw videos on youtube, they are too advanced. Great work for doing all this effort to teach ordinary guys like us.

Preetham
February 17, 2016 at 5:30 am

Hello,
Good introduction into machine learning and SVM.
I had a doubt, the hyperplane which you are referring, does it need to be a straight line or can it be a curve also?

**Alexandre KOWALCZYK** Post author
February 17, 2016 at 10:31 pm

In two dimensions an hyperplane is a straight line, sometimes you can see pictures with a circle for instance but it is just a projection of an hyperplane of a higher dimension into the 2 dimensions. In 3 dimensions it is a plane. In more than 3 dimensions you cannot visualize it.

inesda
February 26, 2016 at 11:45 am

hello your tutorial is excllent ! the title of my thesis is the contextual discovery of web services using svm but the problem I searched and I have never found this with svm I wonder if we can implement web services on svm

**Alexandre KOWALCZYK** Post author
February 27, 2016 at 12:45 am

Thank you. I don't understand what you mean by "implementing web services" on svm. Sorry.

mani
March 2, 2016 at 7:42 am

Great post, keep it up!

Pingback: 机器学习相关网络资源 | 研究生主页

zakaria
March 15, 2016 at 4:47 pm

Please how i do for classify image whit SVM ??

thesoul
March 16, 2016 at 8:01 am

Wow. Well explained. I was so confused with other sites explaining SVM when I stumbled upon your site. Thanks for the simple yet powerful explanation.

Kishor Bhosale
March 18, 2016 at 6:19 pm

Really very simple explanation very easy to understand... I like your work.

abha
April 2, 2016 at 9:34 am

very well explained

Ahmad
April 4, 2016 at 3:04 pm

thanks a lot for this tutorial , I need your help about my problem please

for needle trajectory detection in ultrasound , if I want to estimate the needle trajectory , by classify each pixel in the image to needle class , and background class , if I apply first the log gabor filter what will be the next classifier , can I use svm and how , I use matlab

thank you very much

---

Pradip Nichite
May 18, 2016 at 9:25 pm

Great Explanation !!

---

kamal
May 20, 2016 at 6:48 pm

Hello Alex,
Great article for beginners, I see the explaining starts by picking up the support vectors upfront. Basically those vectors that fall on the decision boundaries are picked up upfront. As I was new to SVM, I was wondering how would a machine do this. What is the mechanism used, is the euclidean distance between two points (vectors) the key?

-Kamal.

---

**Alexandre KOWALCZYK** Post author
May 21, 2016 at 1:41 pm

The support vectors are the ones for which lagrangian multipliers are non zero. This will be explained in the upcoming article about optimization.

---

Sonali Nimkar
May 26, 2016 at 7:01 am

Beautifully explained, thank you very much Alexandre.

Sidra Sultan
May 31, 2016 at 5:55 pm

One of the good articles. Thankyou so much

Srinivasan Aruchamy
June 7, 2016 at 3:27 pm

very Nice way of explanation. Thanks Alexandre

Ayush varshney
July 8, 2016 at 2:26 pm

The best article i've found on this subject.

Thanks for sharing your expert knowledge on the subject, i am really impressed and wanted to learn more from your blogs..

Ngọc Tân Đỗ
August 16, 2016 at 5:32 pm

I like your writing style so much. Thank you very much!

Luiz de Souza
August 20, 2016 at 8:57 am

Very good explanation Simple and didactic. Hard to find combination.

Rohit Tanwar
September 3, 2016 at 6:34 am

I read so many articles on SVM but the clarity of concepts I got from here is awesome.... hope to get your guidance in this area in future too..

robert d
September 15, 2016 at 3:33 pm

This is the best explanation of the maths behind svms i have read and i have read quite a few. Will you include kkt conditions when dealing with optimisation with inequalities ?
Thank you very much for this explanation and i look forward to the book.

**Alexandre KOWALCZYK** Post author
September 15, 2016 at 5:58 pm

Hello Robert,
Yes KKT will be included of course.
Thank you for the kind comment.

polosepian
September 16, 2016 at 1:46 pm

Excellent job

Sanidhya Singh

September 18, 2016 at 8:15 pm

Great explanation in simple way.. Thanks Alexandre. Looking forward to more of your work. Do you have any other tutorials website or blog on other topics. Or if you can mail me also the docs. It will be of great help.

---

**Alexandre KOWALCZYK** Post author

September 19, 2016 at 12:20 am

Thank you very much. No, I do not have other tutorials I am fully focused on SVM 🙂 Maybe later. Which docs are you talking about?

---

Ravindra M

September 19, 2016 at 7:02 am

I wish you could do tutorials for other machine learning techniques as well. Its the best tutorial available in SVM.

---

**Alexandre KOWALCZYK** Post author

September 19, 2016 at 10:22 pm

Thanks a lot 🙂

---

Ganecian

September 24, 2016 at 12:03 am

Nice post... But you're working with continuous dataset in your example so we can easily map it to 2 dimensional vector space. How about discrete/categorized attribute such as skin color (white, black, brown, etc), how to map it?

**Alexandre KOWALCZYK** Post author
September 24, 2016 at 3:21 pm

Usually, you need to one-hot-encode categorical variables. You can read more about preprocessing the data in this article.

Amal Targhi
October 12, 2016 at 10:41 am

thank you it one of the best tutorials

Ming
October 15, 2016 at 2:50 pm

Passionate to work as a data scientist, I've been looking for tutorials giving me ideas what exactly SVM is and this page is the one, Thank you.

**Alexandre KOWALCZYK** Post author
October 16, 2016 at 11:38 pm

Thank you for your kind words. 🙂

jayadi.kurniawan
October 31, 2016 at 1:08 pm

Nice article
I have a question, in your example the SVM used to classify 2 classes
can we use SVM to classify 3 classes ?
my research topic is sentiment analysis (tex classification) using SVM. Im about to classify 3 classes (positive, negative and netral class).
thanks before.

**Alexandre KOWALCZYK** Post author
October 31, 2016 at 6:08 pm

Yes, SVMs can be used to classify more than one class. There are several ways to do this, one-vs-one, one-vs-all, ... You can read this page on the subject if you are using scikit-learn. In my upcoming ebook about SVM there will be one chapter dedicated to multi-class classification as it is a frequent question.

Sonu
November 7, 2016 at 5:51 am

Dear sir,
How to calculate the weight vector w and the bias term b?. I have searched a lot but did not find the clear answer of calculation of the weight vector w and the b. Could you please explain in simple way as you done all the things

**Alexandre KOWALCZYK** Post author
November 15, 2016 at 10:20 pm

Hello Sonu. Computing w and b is done once we have solved the optimization problem and found the Lagrange multipliers. Moreover, some algorithms such as SMO compute w and b separately. I explain how to do that in detail in my upcoming ebook. As it is not published yet, I can recommend you to read this paper by Andrew Ng. You will find how to compute w in equation (9) and b in equation (11).

Ganesh S.

December 5, 2016 at 3:10 pm

Hi,

This is awesome post. It clears every mathematical aspect of support vector machine.
Can you suggest some sources/book for such mathematical explanation of other algorithms ?

Thank you !

---

**Alexandre KOWALCZYK** Post author
December 16, 2016 at 1:40 pm

Hi,
The book Pattern Recognition and Machine Learning by Bishop is very interesting. If you want a much broader view of AI I recommend Artificial Intelligence a Modern Approach by Russel and Norvig.

---

Srinidhi
January 4, 2017 at 3:56 pm

Hi Alexandre,

Thanks for such a lucid presentation to beginners of Machine learning algorithms.Helped me gain confidence that ML is not an bewildering field.

---

Remalli Vidyasagar
February 3, 2017 at 8:31 am

Hi,

it is a very simple tutorail and very clear on SVM. I very much impressed with your explanation.
thank you

Shravan Kumar
February 7, 2017 at 6:06 am

Good Writeup!

wert
March 11, 2017 at 3:35 am

your blog is too good for a beginner like me

ramamurthi kumar
April 2, 2017 at 9:20 am

hi alexandre,

an amazing work, by you on SVM, i have been trying to understand svm for almost a month, and your site filled that void.

Just wanna ask you, can you discuss preceptron, neural networks as well ? or can you suggest any websites / book for the above ?

thanks.

**Alexandre KOWALCZYK** Post author
April 3, 2017 at 11:35 pm

Hello. I discuss perceptron in my upcoming ebook 🙂 Feel free to subscribe to receive a mail when it is ready!

Anil Chaurasiya
April 30, 2017 at 8:50 am

Hey! Alexandre, The way you wrote this topic help lot.
thanks for wrighting such a goood topics.

---

phucdk
May 10, 2017 at 6:51 am

that's great article. wait to read your book.

---

Morgan
May 17, 2017 at 4:28 pm

Hello Alexander, I enjoyed this article and it is making the concept of SVM very clear to me. Please how do I use SVM and fuzzy logic to develop a medical diagnostic system.
Thanks

---

Puneet Jaiswal
May 18, 2017 at 1:24 pm

Thanks. You so much. Great Work. its quite easy to understand.

---

vamshi polapally
June 7, 2017 at 8:37 am

How i can apply svm algorithm to implement automatic identification and data capture using RFID ??

---

Pingback: Support Vector Machine Classification using Raw Python | James D. McCaffrey

Ekki rinaldi
June 13, 2017 at 2:09 am

Can you add one part about how the kernel works? like how rbf or polynomial implemented mathematically

Gleen A. Dalaorao
August 3, 2017 at 12:44 am

Hello Sir,

I love this blog, it makes our SVM life very easy... I'm working on SVM research but I'm new to this, my research is that I want to tweak or make just a simple modification on SVM on it weakness or disadvantage. I hope you can guide me where to focus or what part of SVM will I modify, the important there in is that it is new or novel so that I gather articles on that and start reading. Hope you can share your thoughts and guide. Thanks and more power

Sudhakar Surapaneni
July 13, 2017 at 4:57 pm

Dear Alexandre KOWALCZYK,
I started taking a course on Machine Learning. Found it easy going until my encounter with SVM. Had to go back and forth multiple times to understand the math and gain some intuition. University courses videos helped to some extent and there too got stuck at few places. Your explanation helped to gain further insights and cleared some of the logic I could not get a hang of. I am grateful to you for your contribution.

Regards
Sudhakar

oymj
July 15, 2017 at 5:58 am

I think I can not subscribe to the list by the mail, could you provide another way to get the coming ebook? thanks!

**Alexandre KOWALCZYK** Post author
July 23, 2017 at 9:12 pm

Hello. I can confirm that you have successfully subscribed.

Ansar
July 23, 2017 at 3:25 pm

Can you please explain mathematics involved in identifying support vectors.

**Alexandre KOWALCZYK** Post author
July 23, 2017 at 9:17 pm

See this answer for a good explanation.

Emmanuel Caster
August 17, 2017 at 8:34 pm

Great Job

ENNAJIH Yassin
September 25, 2017 at 5:49 am

Hello, Great Introduction, couldn't have asked for a better one !!

Can you please add a paragraph or two about the regression side of SVM ?
because this explanation mainly focuses on the classification side of SVM… and I am left wondering how does Support Vector Regression (SVR) work? also by finding the optimal hyperplane that minimizes the margin ? if so? how?

Thanks in advance

Adnan
December 12, 2017 at 12:06 pm

Awesome article sir
10000000000 likes and thumbs up from my side

Emad Hejazian
January 4, 2018 at 9:26 am

Thanks a lot
I found out what SVM is today

Ratna
March 8, 2018 at 2:56 pm

Very Nice explanation on SVM. Eager to receive tutorial on Multiclass SVM.

Yasin
March 27, 2018 at 5:49 am

Good writings

Yasin
March 27, 2018 at 5:50 am

Any such writing on multiclass SVM?

---

**Alexandre KOWALCZYK** Post author
March 27, 2018 at 7:23 pm

There is a chapter about it in my free ebook 🙂

---

Tu
April 23, 2018 at 6:21 am

Great explains. Simple, easy to understanding.

---