

SVM Tutorial

SVM - Understanding the math - the optimal hyperplane

This is the **Part 3** of my series of tutorials about the math behind Support Vector Machine.

If you did not read the previous articles, you might want to start the serie at the beginning by reading this article: [an overview of Support Vector Machine](#).

What is this article about?

The main focus of this article is to show you the reasoning allowing us to select the optimal hyperplane.

Here is a quick summary of what we will see:

- How can we find the optimal hyperplane ?
- How do we calculate the distance between two hyperplanes ?
- What is the SVM optimization problem ?

How to find the optimal hyperplane ?

At the end of Part 2 we computed the distance $\|p\|$ between a point A and a hyperplane. We then computed the margin which was equal to $2\|p\|$.

However, even if it did quite a good job at separating the data it was not the *optimal* hyperplane.

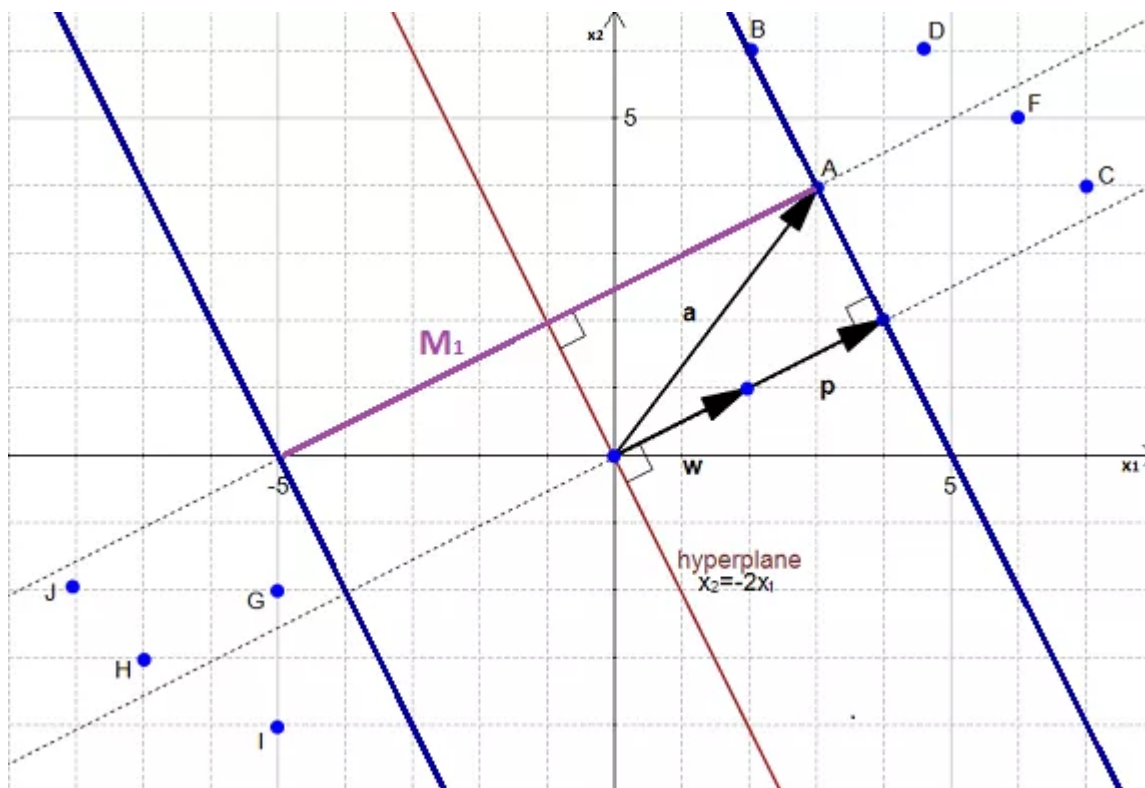


Figure 1: The margin we calculated in Part 2 is shown as M_1

As we saw in Part 1, the optimal hyperplane is the one which maximizes the margin of the training data.

In Figure 1, we can see that the margin M_1 , delimited by the two blue lines, is not the biggest margin separating perfectly the data. The biggest margin is the margin M_2 shown in Figure 2 below.

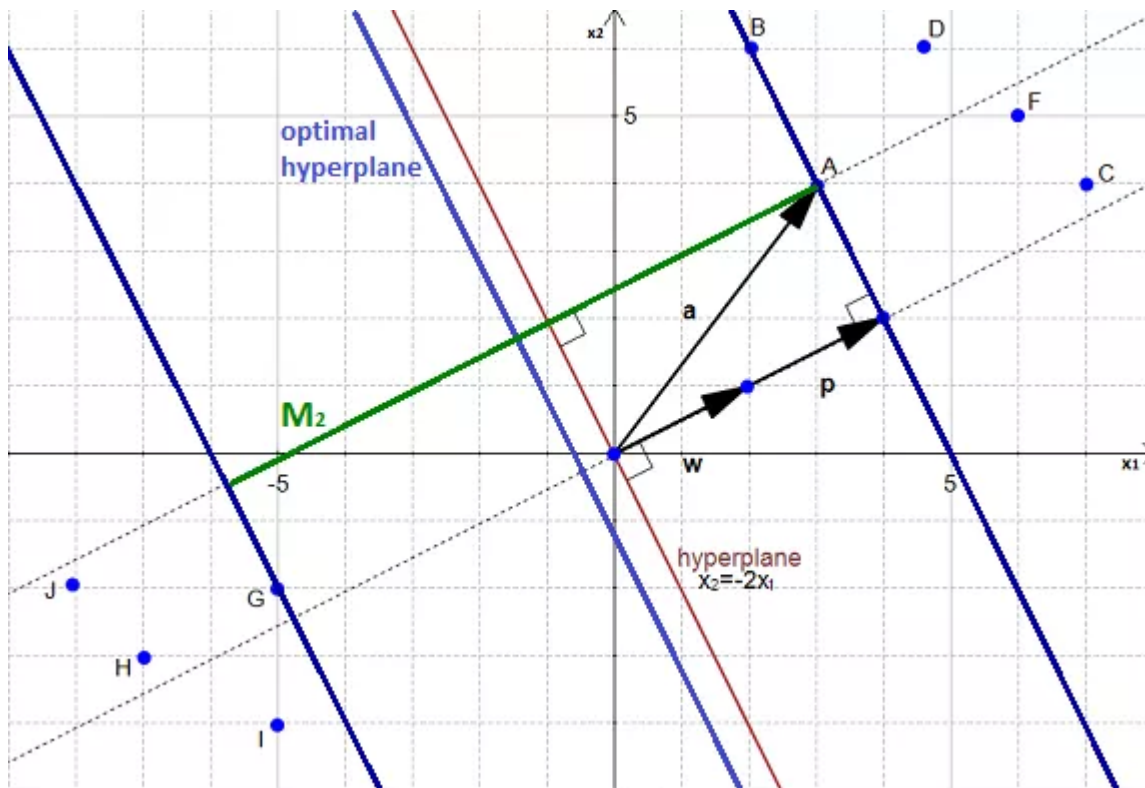


Figure 2: The optimal hyperplane is slightly on the left of the one we used in Part 2.

You can also see the optimal hyperplane on *Figure 2*. It is slightly on the left of our initial hyperplane. How did I find it ? I simply traced a line crossing M_2 in its middle.

Right now you should have the feeling that hyperplanes and margins are closely related. And you would be right!

If I have an hyperplane I can compute its margin with respect to some data point. If I have a margin delimited by two hyperplanes (the dark blue lines in *Figure 2*), I can find a third hyperplane passing right in the middle of the margin.

Finding the biggest margin, is the same thing as finding the optimal hyperplane.

How can we find the biggest margin ?

It is rather simple:

1. You have a dataset
2. select two hyperplanes which separate the data with no points between them
3. maximize their distance (the margin)

The region bounded by the two hyperplanes will be the biggest possible margin.

If it is so simple why does everybody have so much pain understanding SVM ?

It is because as always the simplicity requires some abstraction and mathematical terminology to be well understood.

So we will now go through this recipe step by step:

Step 1: You have a dataset \mathcal{D} and you want to classify it

Most of the time your data will be composed of n vectors \mathbf{x}_i .

Each \mathbf{x}_i will also be associated with a value y_i indicating if the element belongs to the class (+1) or not (-1).

Note that y_i can only have two possible values -1 or +1.

Moreover, most of the time, [for instance when you do text classification](#), your vector \mathbf{x}_i ends up having a lot of dimensions. We can say that \mathbf{x}_i is a p -dimensional vector if it has p dimensions.

So your dataset \mathcal{D} is the set of n couples of element (\mathbf{x}_i, y_i)

The more formal definition of an initial dataset in set theory is :

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p, y_i \in \{-1, 1\}\}_{i=1}^n$$

Step 2: You need to select two hyperplanes separating the data with no points between them

Finding two hyperplanes separating some data is easy when you have a pencil and a paper. But with some p -dimensional data it becomes more difficult because you can't draw it.

Moreover, even if your data is only 2-dimensional it might not be possible to find a separating hyperplane !

*You can only do that if your data **is** linearly separable*

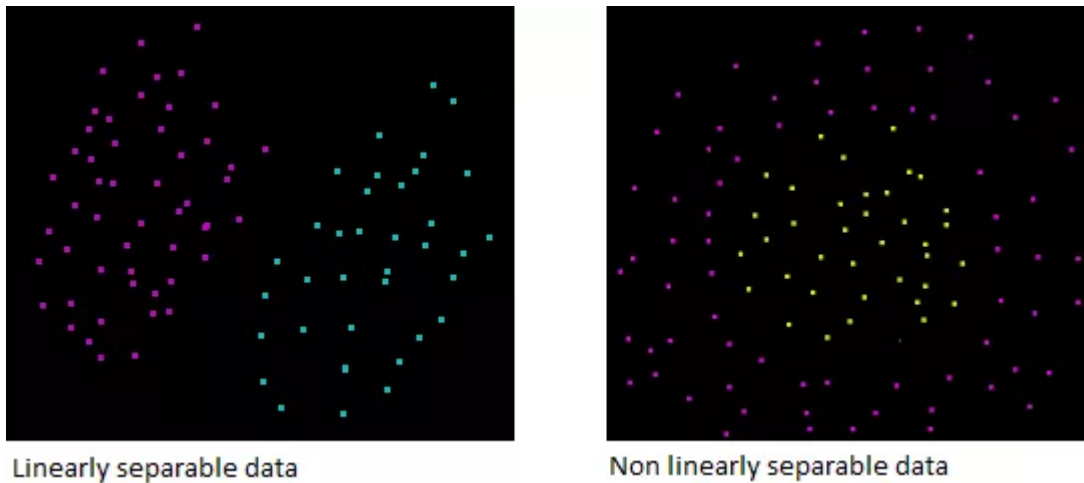


Figure 3: Data on the left can be separated by an hyperplane, while data on the right can't

So let's assume that our dataset \mathcal{D} IS linearly separable. We now want to find two hyperplanes with no points between them, but we don't have a way to visualize them.

What do we know about hyperplanes that could help us ?

Taking another look at the hyperplane equation

We saw previously, that [the equation of a hyperplane](#) can be written

$$\mathbf{w}^T \mathbf{x} = 0$$

However, in the [Wikipedia article about Support Vector Machine](#) it is said that :

Any hyperplane can be written as the set of points \mathbf{x} satisfying $\mathbf{w} \cdot \mathbf{x} + b = 0$.

First, we recognize another notation for the dot product, the article uses $\mathbf{w} \cdot \mathbf{x}$ instead of $\mathbf{w}^T \mathbf{x}$.

You might wonder... Where does the $+b$ comes from ? Is our previous definition incorrect ?

Not quite. Once again it is a question of notation. In our definition the vectors \mathbf{w} and \mathbf{x} have three dimensions, while in the Wikipedia definition they have two dimensions:

Given two 3-dimensional vectors $\mathbf{w}(b, -a, 1)$ and $\mathbf{x}(1, x, y)$

$$\mathbf{w} \cdot \mathbf{x} = b \times (1) + (-a) \times x + 1 \times y$$

$$\mathbf{w} \cdot \mathbf{x} = y - ax + b \quad (1)$$

Given two 2-dimensional vectors $\mathbf{w}'(-a, 1)$ and $\mathbf{x}'(x, y)$

$$\mathbf{w}' \cdot \mathbf{x}' = (-a) \times x + 1 \times y$$

$$\mathbf{w}' \cdot \mathbf{x}' = y - ax \quad (2)$$

Now if we add b on both side of the equation (2) we got :

$$\mathbf{w}' \cdot \mathbf{x}' + b = y - ax + b$$

$$\mathbf{w}' \cdot \mathbf{x}' + b = \mathbf{w} \cdot \mathbf{x} \quad (3)$$

For the rest of this article we will use 2-dimensional vectors (as in equation (2)).

Given a hyperplane H_0 separating the dataset and satisfying:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

We can select two others hyperplanes H_1 and H_2 which also separate the data and have the following equations :

$$\mathbf{w} \cdot \mathbf{x} + b = \delta$$

and

$$\mathbf{w} \cdot \mathbf{x} + b = -\delta$$

so that H_0 is equidistant from H_1 and H_2 .

However, here the variable δ is not necessary. So we can set $\delta = 1$ to simplify the problem.

$$\mathbf{w} \cdot \mathbf{x} + b = 1$$

and

$$\mathbf{w} \cdot \mathbf{x} + b = -1$$

Now we want to be sure that they have no points between them.

We won't select *any* hyperplane, we will only select those who meet the two following **constraints**:

For each vector \mathbf{x}_i either :

$$\mathbf{w} \cdot \mathbf{x}_i + b \geq 1 \text{ for } \mathbf{x}_i \text{ having the class } 1 \quad (4)$$

or

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \text{ for } \mathbf{x}_i \text{ having the class } -1 \quad (5)$$

Understanding the constraints

On the following figures, all red points have the class 1 and all blue points have the class -1 .

So let's look at *Figure 4* below and consider the point *A*. It is red so it has the class 1 and we need to verify it does not violate the constraint $\mathbf{w} \cdot \mathbf{x}_i + b \geq 1$

When $\mathbf{x}_i = A$ we see that the point is on the hyperplane so $\mathbf{w} \cdot \mathbf{x}_i + b = 1$ and the constraint is respected. The same applies for *B*.

When $\mathbf{x}_i = C$ we see that the point is above the hyperplane so $\mathbf{w} \cdot \mathbf{x}_i + b > 1$ and the constraint is respected. The same applies for *D*, *E*, *F* and *G*.

With an analogous reasoning you should find that the second constraint is respected for the class -1 .

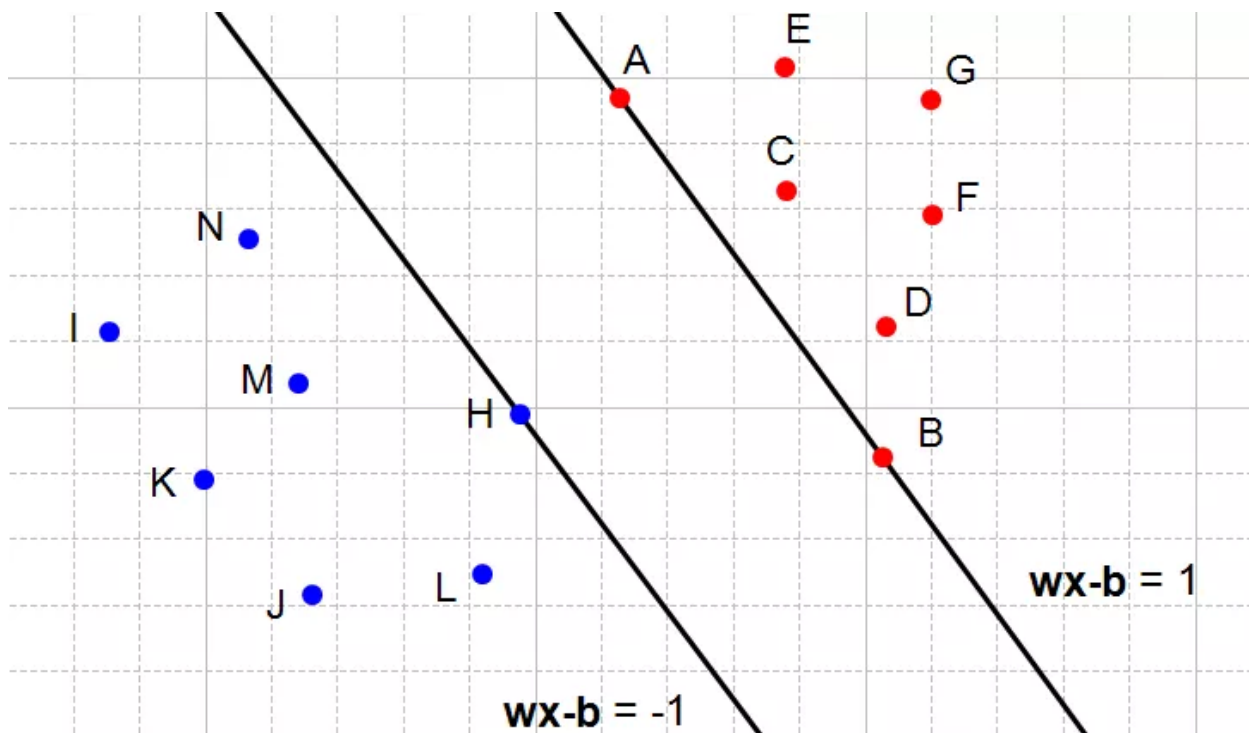


Figure 4: Two hyperplanes satisfying the constraints

On *Figure 5*, we see another couple of hyperplanes respecting the constraints:

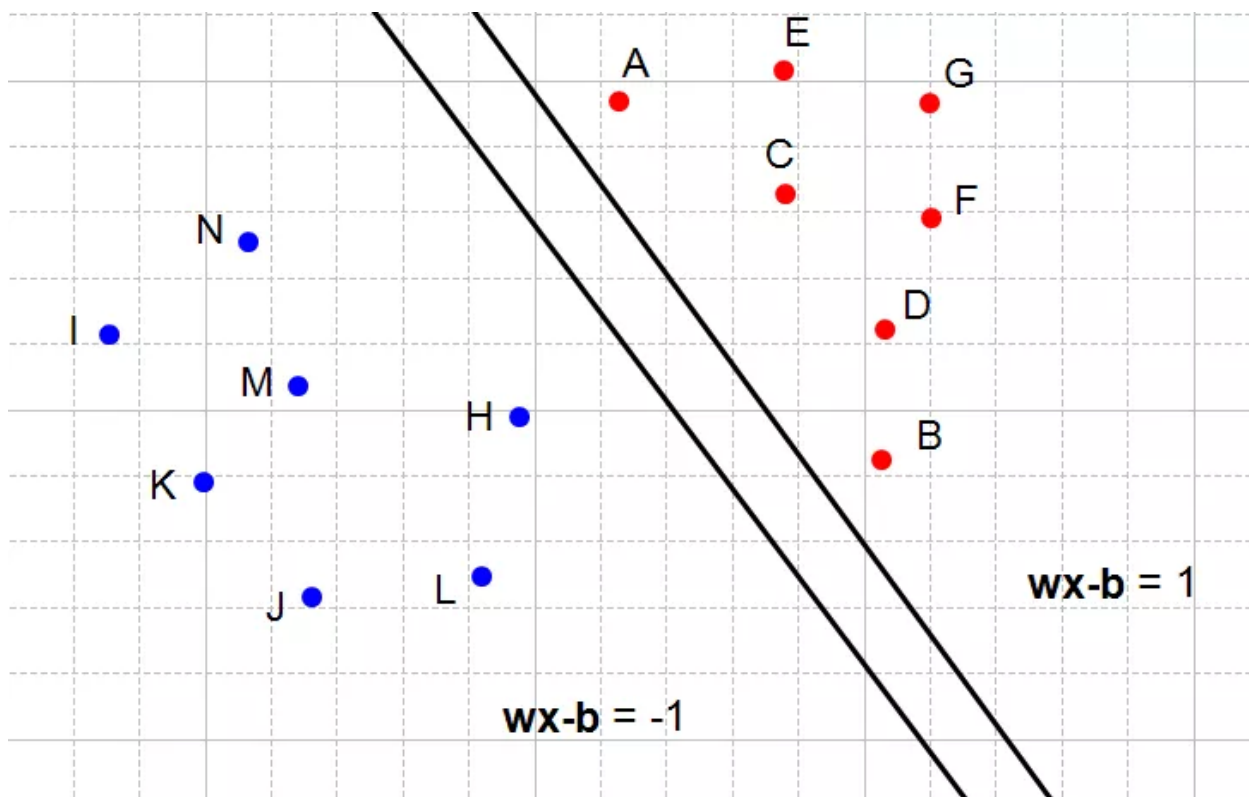


Figure 5: Two hyperplanes also satisfying the constraints

And now we will examine cases where the constraints are not respected:

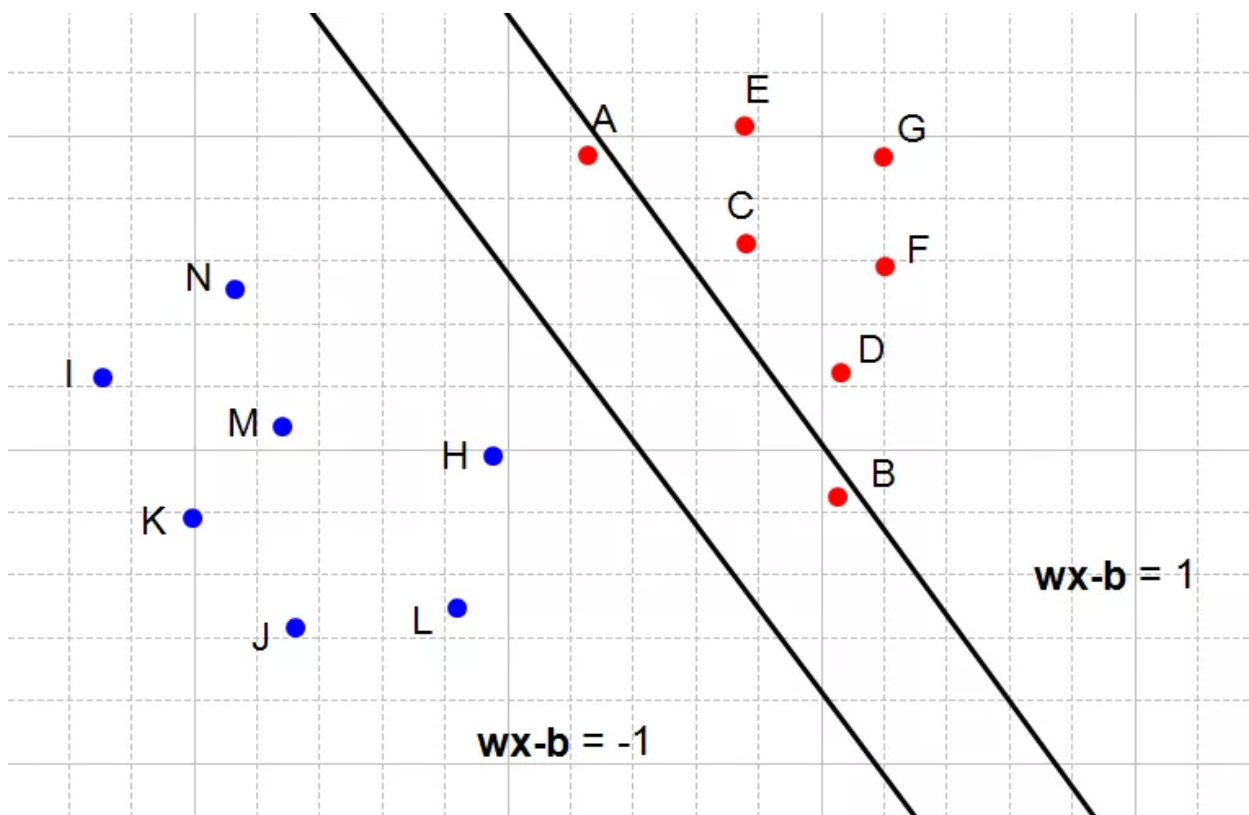


Figure 6: The right hyperplane does not satisfy the first constraint

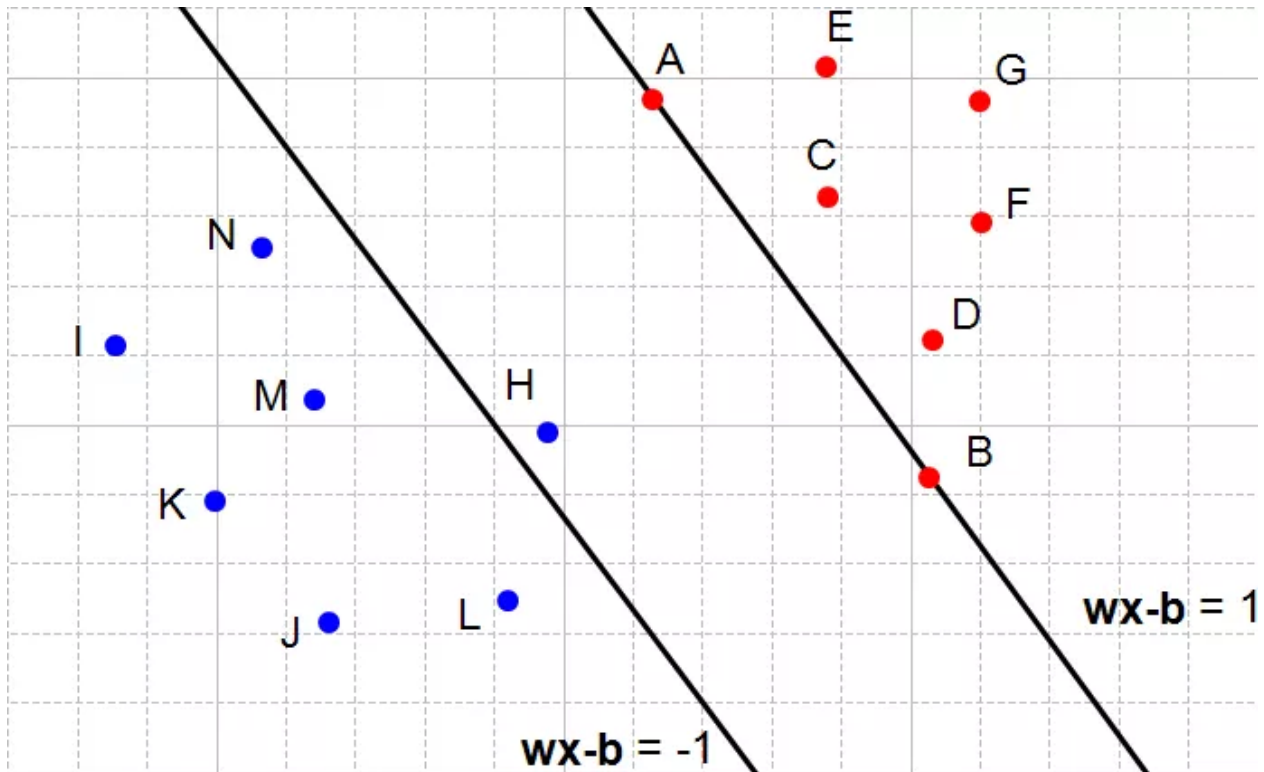


Figure 7: The left hyperplane does not satisfy the second constraint

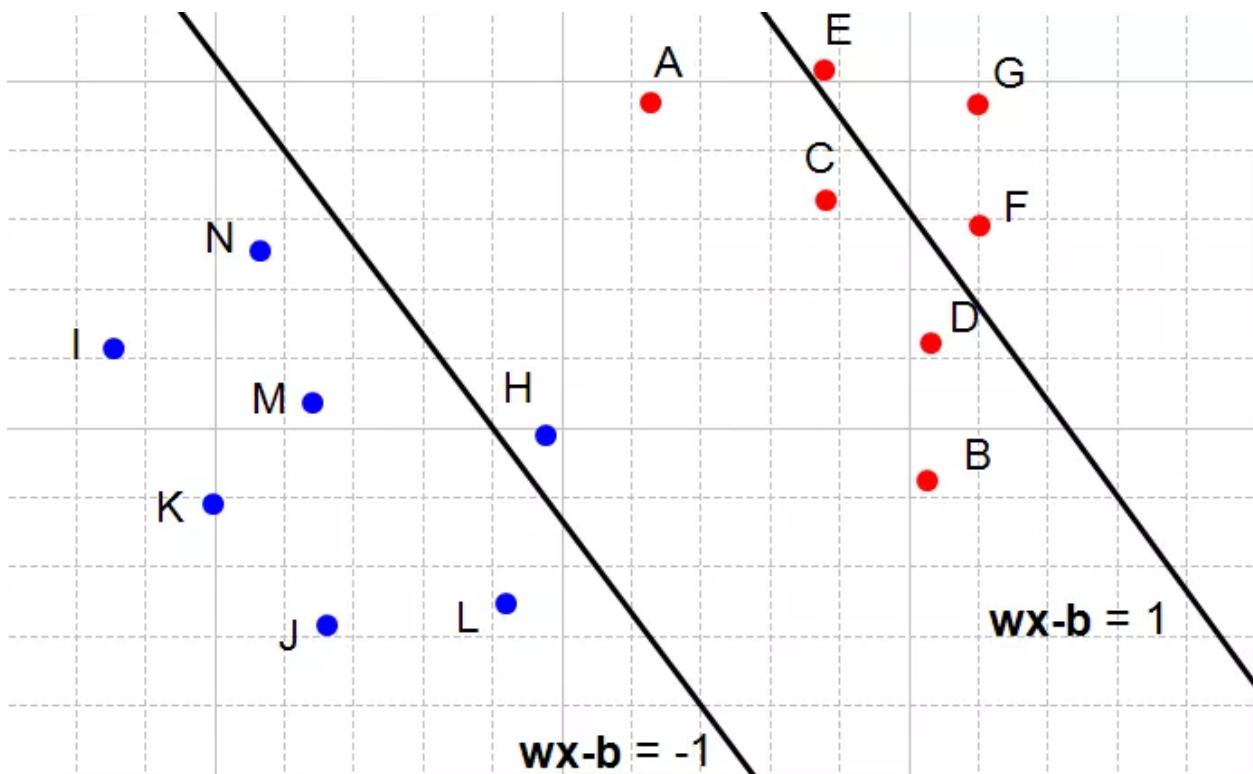


Figure 8: Both constraint are not satisfied

What does it mean when a constraint is not respected? It means that we cannot select these two hyperplanes. You can see that every time the constraints are not satisfied (Figure 6, 7 and 8) there are points

between the two hyperplanes.

By defining these constraints, we found a way to reach our initial goal of selecting two hyperplanes **without points between them**. And it works not only in our examples but also in p -dimensions !

Combining both constraints

In mathematics, people like things to be expressed concisely.

Equations (4) and (5) can be combined into a single constraint:

We start with equation (5)

for \mathbf{x}_i having the class -1

$$\mathbf{w} \cdot \mathbf{x}_i + b \leq -1$$

And multiply both sides by y_i (which is always -1 in this equation)

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq y_i(-1)$$

Which means equation (5) can also be written:

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \text{ for } \mathbf{x}_i \text{ having the class } -1 \quad (6)$$

In equation (4), as $y_i = 1$ it doesn't change the sign of the inequation.

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \text{ for } \mathbf{x}_i \text{ having the class } 1 \quad (7)$$

We combine equations (6) and (7) :

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 \text{ for all } 1 \leq i \leq n \quad (8)$$

We now have a unique constraint (equation 8) instead of two (equations 4 and 5) , but they are mathematically equivalent. So their effect is the same (there will be no points between the two hyperplanes).

Step 3: Maximize the distance between the two hyperplanes

This is probably be the hardest part of the problem. But don't worry, I will explain everything along the way.

a) What is the distance between our two hyperplanes ?

Before trying to maximize the distance between the two hyperplane, we will first ask ourselves: how do we compute it ?

Let:

- \mathcal{H}_0 be the hyperplane having the equation $\mathbf{w} \cdot \mathbf{x} + b = -1$
- \mathcal{H}_1 be the hyperplane having the equation $\mathbf{w} \cdot \mathbf{x} + b = 1$
- \mathbf{x}_0 be a point in the hyperplane \mathcal{H}_0 .

We will call m the perpendicular distance from \mathbf{x}_0 to the hyperplane \mathcal{H}_1 . By definition, m is what we are used to call **the margin**.

As \mathbf{x}_0 is in \mathcal{H}_0 , m is the distance between hyperplanes \mathcal{H}_0 and \mathcal{H}_1 .

We will now try to find the value of m .

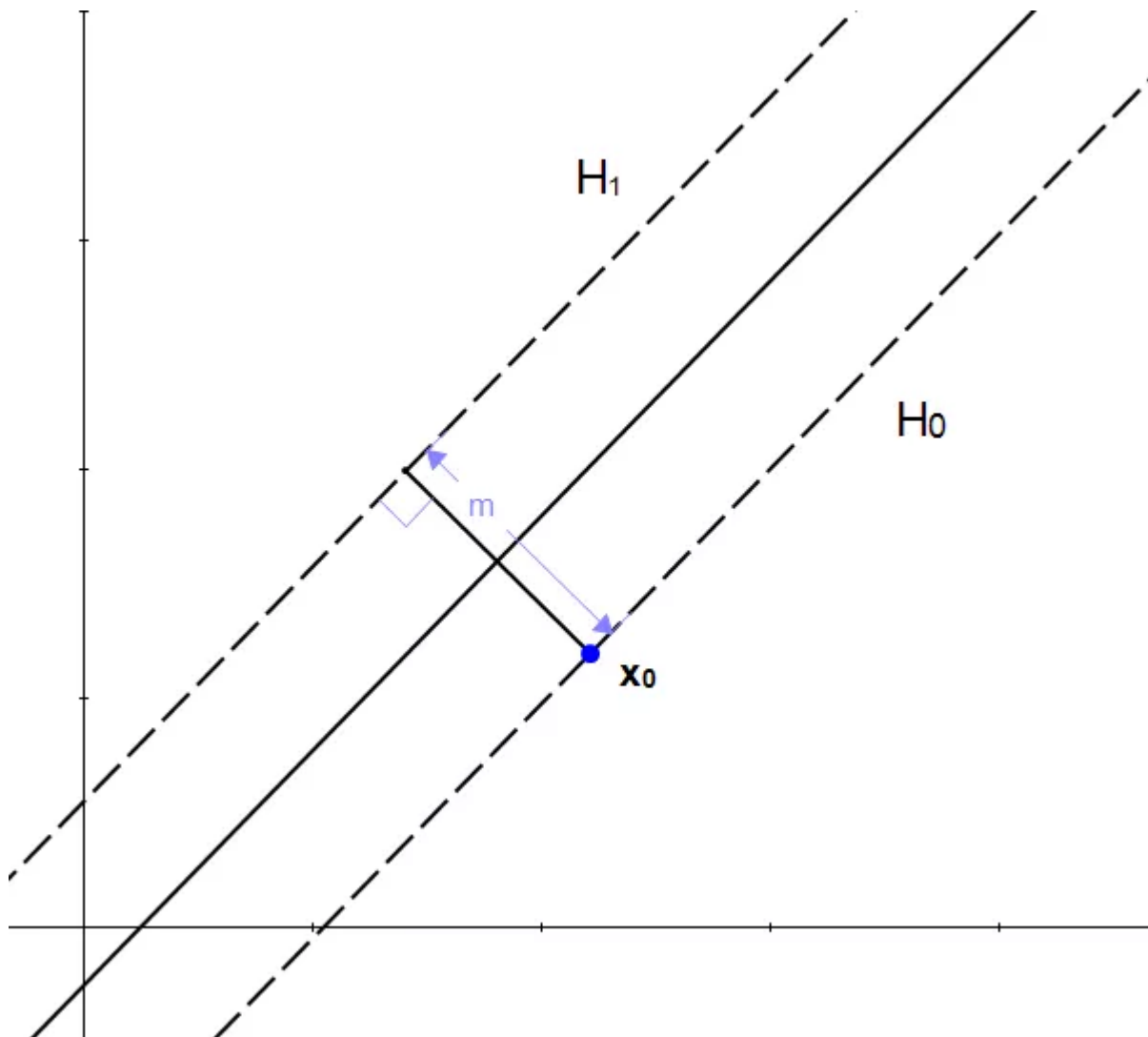


Figure 9: m is the distance between the two hyperplanes

You might be tempted to think that if we add m to \mathbf{x}_0 we will get another point, and this point will be on the other hyperplane !

But **it does not work**, because m is a *scalar*, and \mathbf{x}_0 is a *vector* and adding a scalar with a vector is not possible. However, we know that adding two vectors is possible, so if we transform m into a vector we will be able to do an addition.

We can find the set of all points which are at a distance m from \mathbf{x}_0 . It can be represented as a circle :

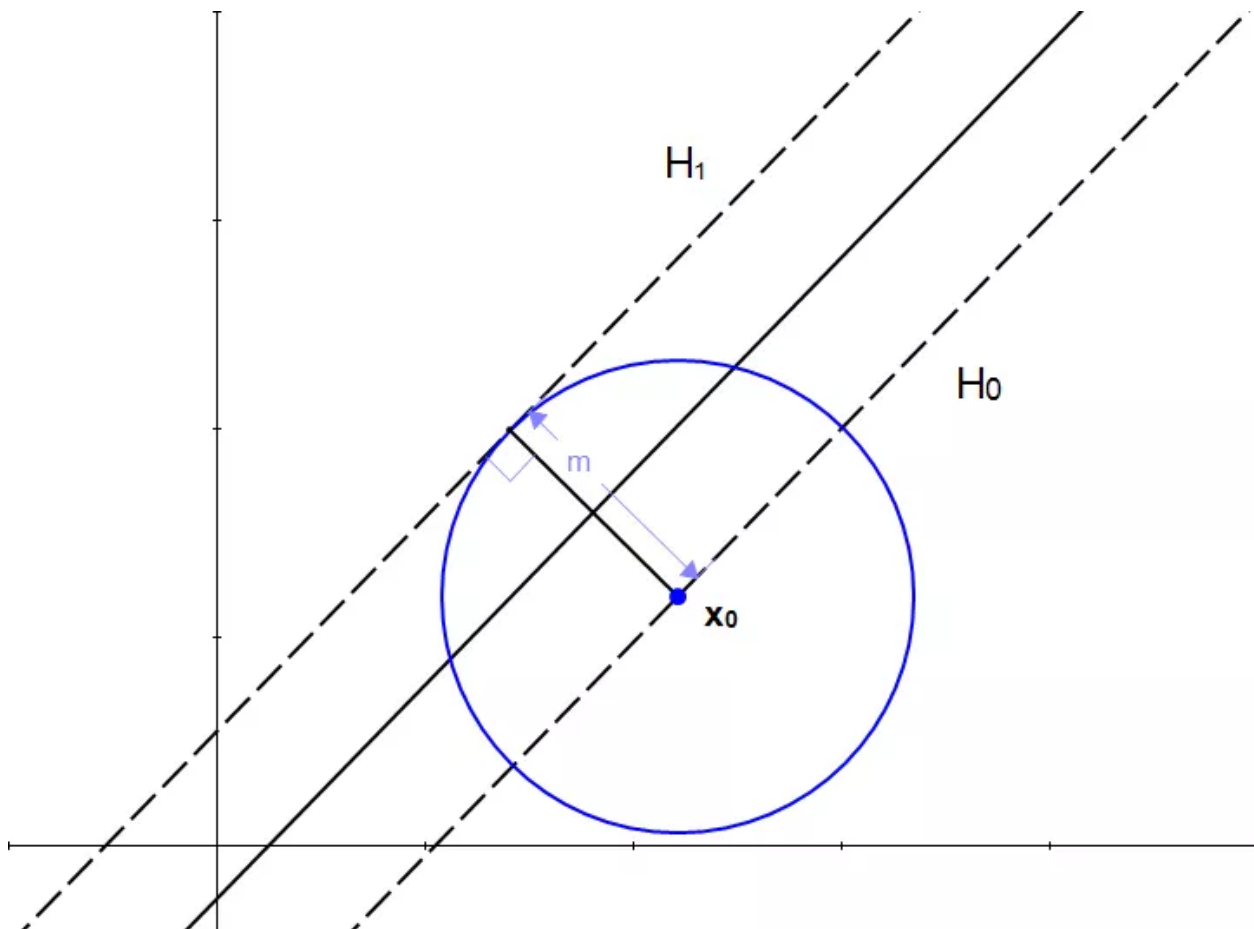


Figure 10: All points on the circle are at the distance m from x_0

Looking at the picture, the necessity of a vector become clear. With just the length m we don't have one crucial information : **the direction**. (recall from Part 2 that a vector has a magnitude and a direction).

We can't add a scalar to a vector, but we know if we multiply a scalar with a vector we will get another vector.

From our initial statement, we want this vector:

1. to have a magnitude of m
2. to be perpendicular to the hyperplane \mathcal{H}_1

Fortunately, we already know a vector perpendicular to \mathcal{H}_1 , that is \mathbf{w} (because $\mathcal{H}_1 = \mathbf{w} \cdot \mathbf{x} + b = 1$)

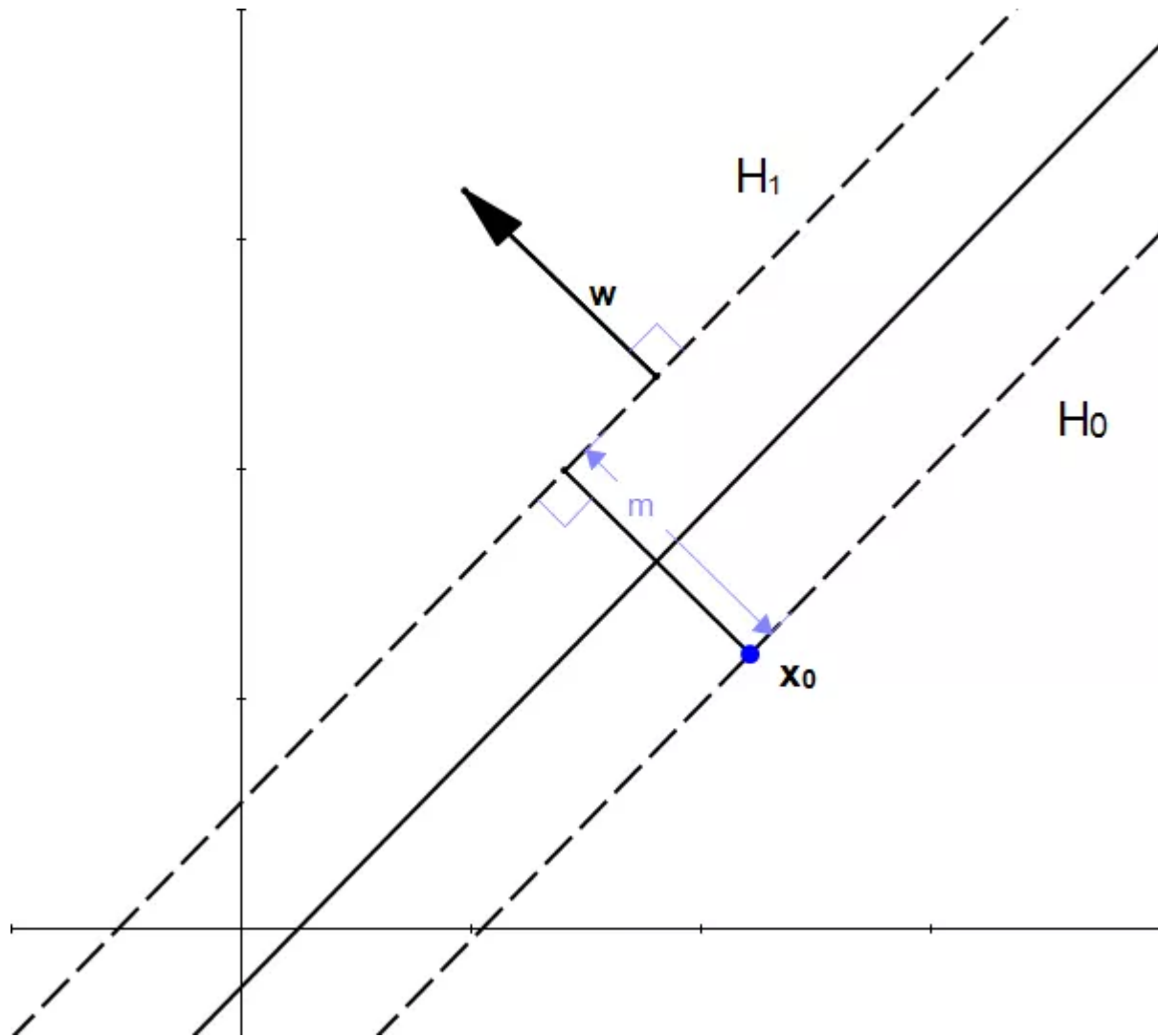


Figure 11: \mathbf{w} is perpendicular to H_1

Let's define $\mathbf{u} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$ the unit vector of \mathbf{w} . As it is a unit vector $\|\mathbf{u}\| = 1$ and it has the same direction as \mathbf{w} so it is also perpendicular to the hyperplane.

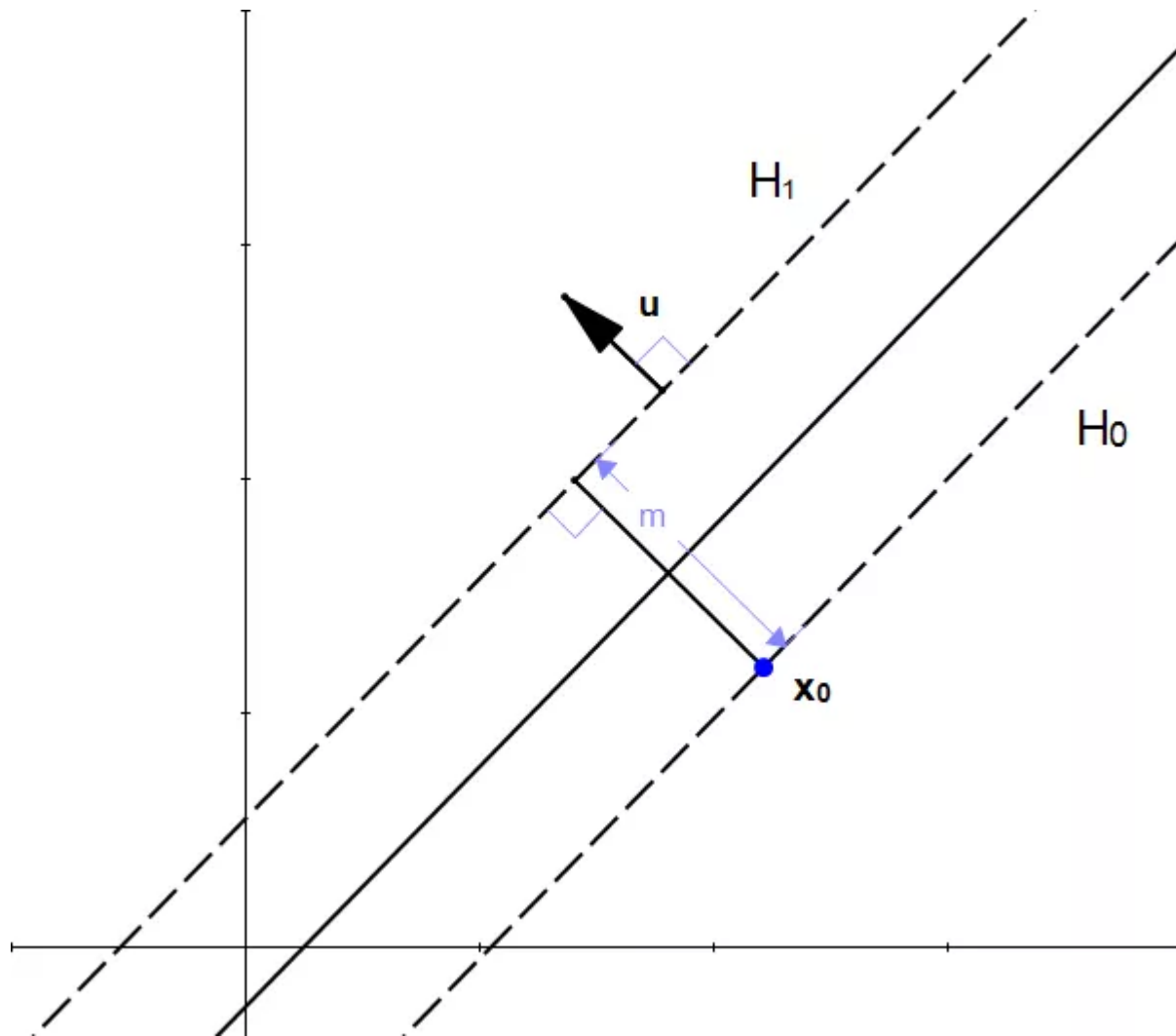


Figure 12: u is also perpendicular to H_1

If we multiply \mathbf{u} by m we get the vector $\mathbf{k} = m\mathbf{u}$ and :

1. $\|\mathbf{k}\| = m$
2. \mathbf{k} is perpendicular to \mathcal{H}_1 (because it has the same direction as \mathbf{u})

From these properties we can see that \mathbf{k} is the vector we were looking for.

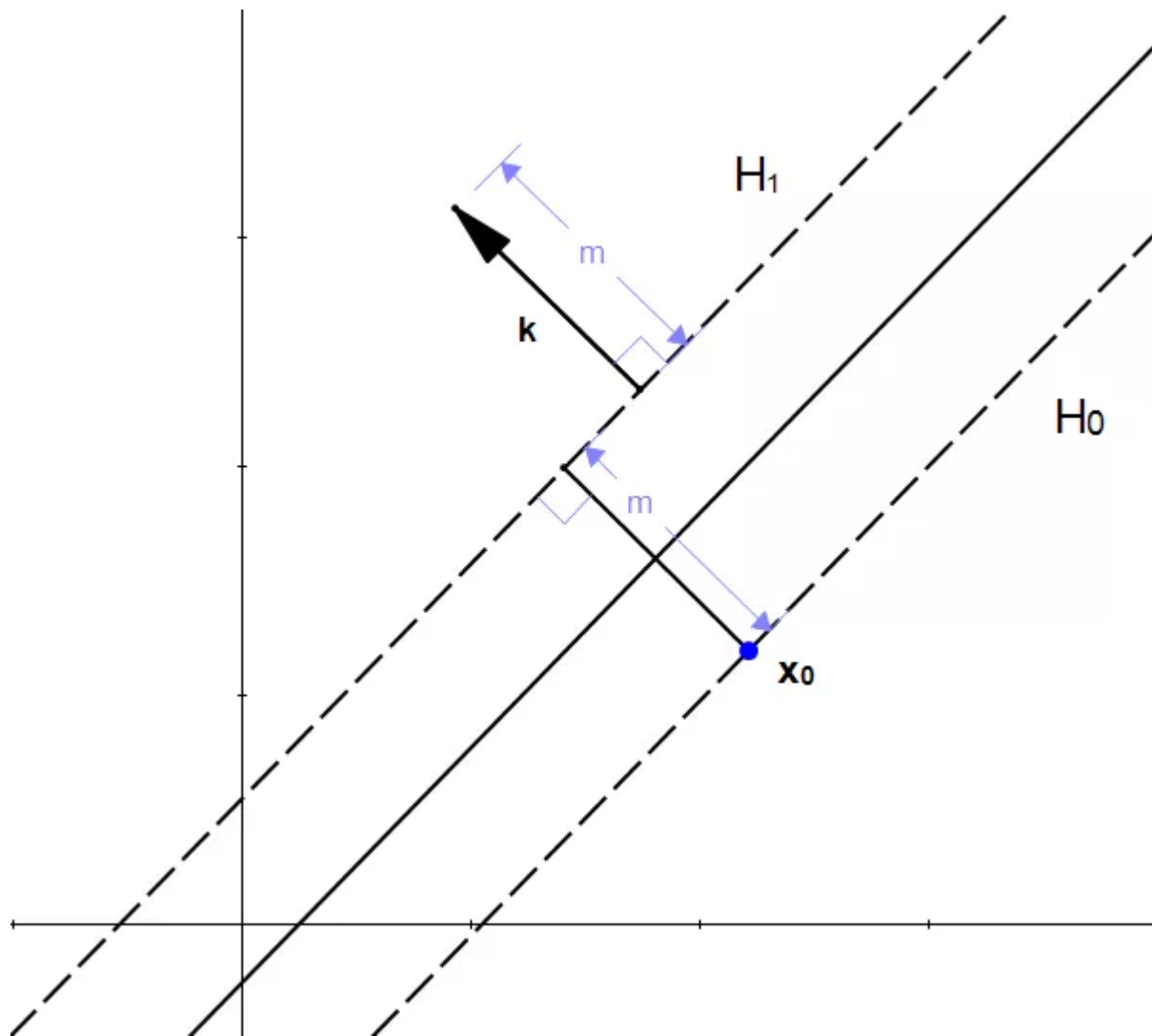


Figure 13: k is a vector of length m perpendicular to H_1

$$\mathbf{k} = m\mathbf{u} = m \frac{\mathbf{w}}{\|\mathbf{w}\|} \quad (9)$$

We did it ! We transformed our scalar m into a vector \mathbf{k} which we can use to perform an addition with the vector \mathbf{x}_0 .

If we start from the point \mathbf{x}_0 and add k we find that the point $\mathbf{z}_0 = \mathbf{x}_0 + \mathbf{k}$ is in the hyperplane \mathcal{H}_1 as shown on Figure 14.

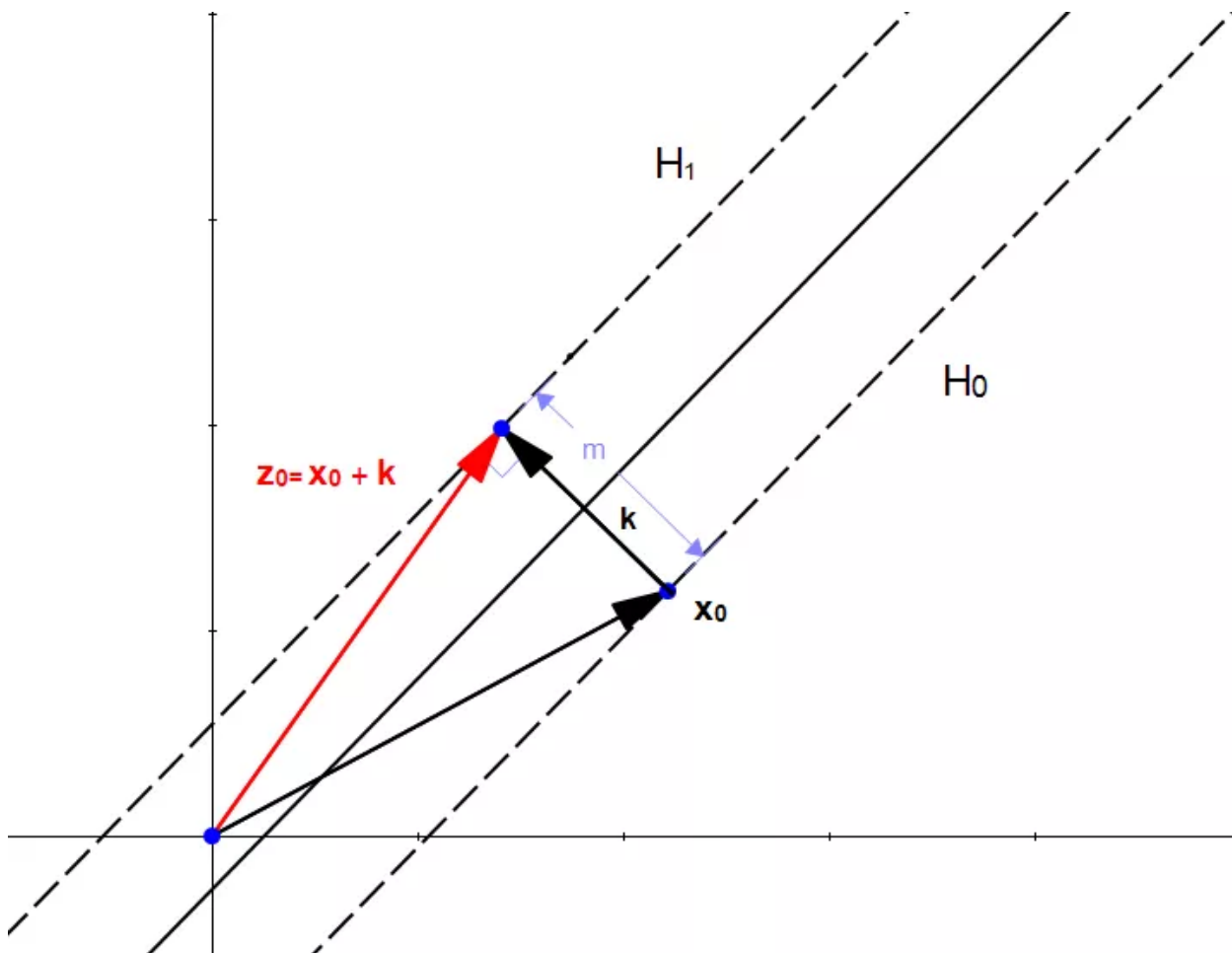


Figure 14: z_0 is a point on H_1

The fact that \mathbf{z}_0 is in \mathcal{H}_1 means that

$$\mathbf{w} \cdot \mathbf{z}_0 + b = 1 \quad (10)$$

We can replace \mathbf{z}_0 by $\mathbf{x}_0 + \mathbf{k}$ because that is how we constructed it.

$$\mathbf{w} \cdot (\mathbf{x}_0 + \mathbf{k}) + b = 1 \quad (11)$$

We can now replace \mathbf{k} using equation (9)

$$\mathbf{w} \cdot \left(\mathbf{x}_0 + m \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + b = 1 \quad (12)$$

We now expand equation (12)

$$\mathbf{w} \cdot \mathbf{x}_0 + m \frac{\mathbf{w} \cdot \mathbf{w}}{\|\mathbf{w}\|} + b = 1 \quad (13)$$

The dot product of a vector with itself is the square of its norm so :

$$\mathbf{w} \cdot \mathbf{x}_0 + m \frac{\|\mathbf{w}\|^2}{\|\mathbf{w}\|} + b = 1 \quad (14)$$

$$\mathbf{w} \cdot \mathbf{x}_0 + m\|\mathbf{w}\| + b = 1 \quad (15)$$

$$\mathbf{w} \cdot \mathbf{x}_0 + b = 1 - m\|\mathbf{w}\| \quad (16)$$

As \mathbf{x}_0 is in \mathcal{H}_0 then $\mathbf{w} \cdot \mathbf{x}_0 + b = -1$

$$-1 = 1 - m\|\mathbf{w}\| \quad (17)$$

$$m\|\mathbf{w}\| = 2 \quad (18)$$

$$m = \frac{2}{\|\mathbf{w}\|} \quad (19)$$

This is it ! We found a way to compute m .

b) How to maximize the distance between our two hyperplanes

We now have a formula to compute the margin:

$$m = \frac{2}{\|\mathbf{w}\|}$$

The only variable we can change in this formula is the norm of \mathbf{w} .

Let's try to give it different values:

When $\|\mathbf{w}\| = 1$ then $m = 2$

When $\|\mathbf{w}\| = 2$ then $m = 1$

When $\|\mathbf{w}\| = 4$ then $m = \frac{1}{2}$

One can easily see that the bigger the norm is, the smaller the margin become.

Maximizing the margin is the same thing as minimizing the norm of \mathbf{w}

Our goal is to maximize the margin. Among all possible hyperplanes meeting the constraints, we will choose the hyperplane with the smallest $\|\mathbf{w}\|$ because it is the one which will have the biggest margin.

This give us the following [optimization problem](#):

Minimize in (\mathbf{w}, b)

$$\|\mathbf{w}\|$$

subject to $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1$

(for any $i = 1, \dots, n$)

Solving this problem is like solving and equation. Once we have solved it, we will have found the couple (\mathbf{w}, b) for which $\|\mathbf{w}\|$ is the smallest possible and the constraints we fixed are met. Which means we will have the equation of the optimal hyperplane !

Conclusion

We discovered that finding the optimal hyperplane requires us to solve an optimization problem. Optimization problems are themselves somewhat tricky. And you need more background information to be able to solve them. So we will go step by step. Let us discover [unconstrained minimization problems in Part 4](#)! Thanks for reading.



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, vectors on June 8, 2015 [<https://www.svm-tutorial.com/2015/06/svm-understanding-math-part-3/>] .

145 thoughts on “SVM - Understanding the math - the optimal hyperplane”



dragon518

June 9, 2015 at 4:47 am

Thank you very much! I have never read a technology blog more fundamental, detailed, clear than yours before. You are a specialist on SVM. BTW, look forward to read part-4.



snake2971chris

June 9, 2015 at 10:57 am

really nice report! Just one thing, if you have data points in a 2 dimensional space(x_1, x_2), like in your examples, you should visualize it in 3 dimensions since the hyperplane is spanned in a third dimension(y). If you don't want to mess around with 3 dimensions use a simple 1D example, so your hyperplane can be really visualized with a line.

That was a thing which confused me for quiet a bit. But once I had the right visualization everything else became kind of intuitive.



Alexandre KOWALCZYK

Post author

June 9, 2015 at 7:31 pm

Yeah sure. Here the third dimension is depicted by the color, but indeed it can be confusing. Thanks for your feedback 😊



Ezra Drieka

March 21, 2018 at 2:09 pm

Thanks so much bro - beginner



sarath r nair

June 17, 2015 at 12:30 pm

Very great post Please dont stop in the middle . Thumbs up.

**Rokesh Jankie**

June 23, 2015 at 6:25 pm

Really enjoying the article. Thanks for the clear description of something very powerful.

**korawit**

July 1, 2015 at 8:47 am

Thanks looking forward to how to solve SVM equation

**Winson**

July 6, 2015 at 7:57 pm

Waiting for part 4. This tutorial is very clear for beginner. 😊 Thanks

**genetixx01**

July 7, 2015 at 10:50 pm

Very helpful. I look forward for part 4.

**Abhinand (Abhi) Sivaprasad**

July 11, 2015 at 9:32 pm

Why are the hyper planes set to 1 and -1. It seems arbitrary? This essentially sets them 2 units apart. What if the data points with labels 1 and labels -1 were very far away?

**Alexandre KOWALCZYK**

Post author

July 13, 2015 at 8:51 pm

They are set to 1 and -1 because it is the only possible values the vector y can take. As suggested by snake2971chris in a previous comment, visualizing the picture in 3 dimensions can help to understand this.

**mlsvm**

July 12, 2015 at 8:37 am

Great SVM series Alexandre! I have a question for this part, specifically on "Fortunately, we already know a vector perpendicular to $\square 1$, that is w (because $\square 1 = w \cdot x - b = 1$)"

Why is w perpendicular to the hyperplane? I've read an explanation over here:

<http://datascience.stackexchange.com/questions/6054/in-svm-algorithm-why-vector-w-is-orthogonal-to-the-separating-hyperplane>

but I still have no clue about the logic behind.

**Alexandre KOWALCZYK**

Post author

July 13, 2015 at 8:59 pm

This is by definition. The equation is called a *normal equation* for the hyperplane. I invite you to read [the part about hyperplanes](#) in this course (page 9). This is about the calculus of functions of several variables but I used it as a reference several times to write this serie of articles (you can read the [full course here](#)).

**archkrish**

October 18, 2015 at 4:34 am

The hyperplanes are defined in the way that there is vector which is perpendicular to all the vectors in the hyperplane...imagine hyperplane consists of all the vectors which are perpendicular to w . There are good video lectures available which explains hyperplane.

**Edemir Ferreira**

August 17, 2015 at 9:40 pm

Thanks for the explanation! I'm anxious for the next part =)

**daniel**

October 11, 2015 at 5:12 pm

simple yet insightful explanation on the math behind the scene...cant wait to see the release of part 4...thanks for your effort in producing this tutorial, very helpful

**Rahul**

October 11, 2015 at 9:18 pm

Hey, really great tutorial, waiting for the next part. I had a doubt...Please help!!
Let us consider a two dimensional dataset as shown in your tutorial. My doubt is what is the physical significance of w . I know it is a vector perpendicular to the hyperplane but what does it actually signify

**Alexandre KOWALCZYK**

Post author

October 12, 2015 at 9:39 pm

I am afraid I don't understand your question. What do you mean by physical significance?

**Nrupatunga**

October 13, 2015 at 4:02 pm

Thank you so much for such a nice and easy tutorial.

**Tejas Khot**

November 1, 2015 at 3:29 pm

Hello Alexandre!

Thank you for an excellent series of posts.

By when can we expect the next part to be released ?

**Alexandre KOWALCZYK**

Post author

November 1, 2015 at 11:43 pm

Hi Tejas. Thanks for your comment. I don't have a precise date because I have been very busy at work. As there is a lot of people asking me I will try to do my best to publish it soon (before 2016)

I added a small form at the end of the article so that you can put your email in it if you want me to send you a mail when it will be published.

**Hugo**

November 2, 2015 at 5:08 am

Thanks for the great tutorial. The expressions $wx-b=1$ and $wx-b=-1$ are however a little confusing. You write the following in defence of these equations.

"And we choose +1 and -1 because they are the only values the class can take."

Question: My understanding is that the result of $wx-b$ should not evaluate to a class label. What would evaluate to a class label is an expression such as $f(wx-b)$ where f is the classification rule that assigns the product between a given test vector and the w vector to the classes 1 or -1. Can you clarify why you are having $wx-b$ evaluating to the value of the class labels? Some literature suggests that the numerical values of the class labels are only arbitrary -- in fact -1 and 1 are chosen in SVM algebra because they enable an elegant way to represent the SVM (e.g., we are able to write $y(wx-b)>0$ because y can only be 1 or -1).

Could you have chosen 1 and -1 because you wanted to have the planes $wx-b=-1, 1$ equidistant from $wx-b=0$. In that case does it mean that I could have used numbers such as 2 and -2?

**Alexandre KOWALCZYK**

Post author

November 4, 2015 at 11:15 pm

Thanks for your comment Hugo.

Indeed the expression "And we choose +1 and -1 because they are the only values the class can take." was an error. I checked online and this value has no connection with the class label.

You can [see equation \(2\) of this paper for more details](#).

I updated the article so that it is more explicit how this number is chosen.

I think we could use numbers such as 2 and -2 but I am not sure, I need to check the math to be 100% sure.

Thanks a lot.

**seokhyun**

March 24, 2016 at 10:43 am

Thanks for the grate tutorial!. To me it was hard to understand, but now i come to realize!

**ZC Liu**

August 4, 2016 at 9:15 am

Actually, you can choose +2 and -2 instead of +1 and -1.

Then you follow the steps: 10 ~ 19, you will get $m = 4 / w'$.

means you find a w' , which length is half of the original one(w), when you choose +1 and -1.

then the support hyper plane is: $w' * x + b' = 2 \Rightarrow 2w * x + 2b = 2$, which is equivalent to the original one: $w * x + b = 1$.

so you can choose other number either, eventually, you will get the same support hyper plane.

**Mutaz**

November 6, 2015 at 7:37 pm

Really thanks a lot for this series

This is the first time I see such a simplicity of explaining the SVM

I am waiting part 4, I am a bit scared to learn how solve that optimization problem. Please try to explain it as simple as possible.

Thanks a lot



Alexandre KOWALCZYK Post author

November 8, 2015 at 4:36 pm

Thanks. I will try 😊



강산하

November 25, 2015 at 4:54 am

I really really appreciate for your article.

I am graduate school student in South Korea

I have a question,

Why do you set $\delta = 1$?

I think δ is related to margin and

it should smaller if there are data point and equation (4), (5) are not true.

What do you think?



Alexandre KOWALCZYK Post author

December 18, 2015 at 10:31 pm

We could have kept δ in all the following equations. At the end we would have had $y_i(\mathbf{w} \cdot \mathbf{x}_i - b) \geq \delta$ in equation 8. Indeed, $\delta = y_i(\mathbf{w} \cdot \mathbf{x}_i - b)$ is the functional margin. You can read part 3 of [this course](https://www.svm-tutorial.com/2015/06/svm-understanding-math-part-3/) to

understand better why we can set it to 1 and -1. You will see in part 4 that he replaces the functional margin by 1. I hope this helps you.

**Ganecian**

October 1, 2016 at 2:15 am

what does Δ means? I still didn't understand this paragraph,
"We now note that we have over-parameterized the problem: if we scale w , b and Δ by a constant factor α , the equations for x are still satisfied. To remove this ambiguity we will require that $\Delta = 1$, this sets the scale of the problem, i.e. if we measure distance in millimeters or meters. " Can you explain this?

**Gleen A. Dalaorao**

August 2, 2017 at 4:18 pm

Hi, is this forum still active?

**강산하**

November 25, 2015 at 5:02 am

another question

In step 2, we select 2 hyperplane, so I think w is fixed in that time

but step3 you tell we minimize w to get biggest margin

I think w never change after step2...

Can you explain?

**Alexandre KOWALCZYK**

Post author

December 18, 2015 at 10:09 pm

In step 2 we select two hyperplanes among all possible couple of hyperplanes satisfying the constraints (4) and (5). Given one hyperplane separating the data, we can find two other hyperplanes meeting the constraint. Given another hyperplane separating the data (with a different w than the first one), we can find two other different hyperplanes meeting the constraint. So minimizing w in step 3 allow us to find the two hyperplanes having the biggest margin and the optimal hyperplane.

**Bruce**

November 28, 2015 at 6:27 pm

Thank you very much Alexandre! I was exactly looking for a text like yours.

**Chuan**

November 30, 2015 at 8:06 pm

A really good series of articles! Thank you!

**knight**

December 18, 2015 at 1:52 am

Please carry on with 4th article, the 3 articles are so informative and extremely useful

**Cao Fuxiang**

December 25, 2015 at 11:54 am

Hello Alexandre!

Thank you very much for your excellent series of posts. You convert a very complicated math problem into a very clear and simple one. I have never seen such a great technical blog before!

I look forward to reading your part-4.

Thank you once again.



agarwalayush9

January 5, 2016 at 7:31 am

Thanks, waiting for the next tut.



Reshma

February 2, 2016 at 3:21 pm

Really an awesome tutorial with the unique way of explanations. Each part of SVM was soo well explained. You are an expert in SVM. Looking forward for the part 4.



arkanra

February 6, 2016 at 1:26 am

Thank you so much for very clear articles, I look forward for part 4



Alexandre KOWALCZYK

Post author

February 7, 2016 at 11:55 pm

You are welcome. I am still very busy right now but I don't give up on part 4. I just have some more SVM related work to do before.



Rajive Jain

February 6, 2016 at 1:04 pm

This is the clearest explanation of a complex topic that I have ever seen. Why can't the professors in classes teach the same way? I have a gut feeling that they don't do it because they don't understand it themselves! I went through your entire tutorial in a couple of hours and understood the entire conceptual framework of SVM - something I never got from my professors in the entire semester. I am sure it took you many days of hard work to put this together. Thanks for sharing your knowledge and understanding - I really appreciate your service to the community.

**Alexandre KOWALCZYK**

Post author

February 7, 2016 at 11:54 pm

Thank you very much for such a kind comment.

**jonathanbrophy47**Jonathan Brophy

February 8, 2016 at 6:24 am

This is such a great and clear explanation of SVMs, easily the best I've read, can't wait for the next tutorial!

**tyvannguyen**

February 16, 2016 at 11:49 am

I've to say that this one is one of the best tutorial I've read.

It can viewed as a beginner-to-intermediate lecture with all necessary visual demonstrations.

Thank you very much.

**Vaibhav Daga**

February 20, 2016 at 12:00 pm

nicely written!!

**Wael**

February 20, 2016 at 1:52 pm

Thank You Alexandre, great tutorial made SVM look so simple.

**Manziba**

February 24, 2016 at 12:40 am

Its wonderful article. So simple explanation which I was looking for. Can you please upload more articles explaining Lagrangian to find optimal w

**Alexandre KOWALCZYK**

Post author

February 27, 2016 at 12:44 am

This will be in Part 4.

**Dong**

February 25, 2016 at 3:56 am

Look forward to part 4. This series is amazing

**johnp631**

February 25, 2016 at 4:42 pm

Great tutorial!

Peach



February 25, 2016 at 7:04 pm

Awesome! Look forward to chapter 4



brian

February 29, 2016 at 5:43 am

Excellent tutorial but please make time for more!



Praveen

March 2, 2016 at 1:29 am

I have a question on the dot product of the 2 vectors $w \cdot x - b = 1$ and -1 .

Isn't it be >0 and <0 . Since we are taking the dot product of the normal vector w with another point x which is like dot product of the two vectors

$|w| |x| \cos(\theta)$ and the values will vary between 1 and -1 and when they are on the hyperplane the values of the dot product will be 0.

Please correct me if I am wrong.



Alexandre KOWALCZYK

Post author

March 2, 2016 at 9:20 pm

We don't want the point to be too close to the initial hyperplane, that is why we introduce delta. We choose two hyperplanes at a distance delta from the original hyperplane. We then arbitrarily set $\delta = 1$. But that means that points on the other hyperplanes will have values $+1$ or -1



Vamshi Dhar

March 3, 2016 at 10:47 pm

Dear Praveen,

when they are on the hyperplane the values of the dot product will not be 0, but 1 on one side of plane and it will be >1 as the point goes away from that plane.

Same way with other plane having -1 for point that is exactly on the plane and as it goes out of the plane then it will be < -1 .

I guess this can be your answer. If i understood some thing from the Alexandre KOWALCZYK's 3 valuable "Understanding the math" parts.

Cheers to Alexandre KOWALCZYK



Vamshi Dhar

March 3, 2016 at 10:54 pm

Dear Alexandre KOWALCZYK,
Desperately waiting for your fourth part .



Ulrich Armel

March 12, 2016 at 4:34 am

Really great article, it really made have a good feeling of SVM after I stop learning machine learning when I came across it and it was very poorly explained and I could n't understand anything.



imen

March 16, 2016 at 3:16 pm

I thank you a lot for this great job. It is amazing to find things very simple lik this. you are an angle.



Tonja Rand

March 16, 2016 at 3:57 pm

Hey,

amazing post!

However, I am a bit confused about the definitions of "hyperplane" and "margins". The two hyperplanes you have H_0 and H_1 (dotted lines) are the actual support vectors. A non-dotted line is a separating hyperplane.

Right? I found in the internet two different definitions of "margins". In this figure

(<https://onionesquereality.files.wordpress.com/2009/03/optimal-margin-classifier.jpg>) the margin is defined as a distance between a support vector and a separating hyperplane. In another figure

(<http://blog.pengyifan.com/wp-content/uploads/2013/09/svm.png>) margin is a distance between two support vectors. And as far as I understood you define it also as a distance between two support vectors.

If there are two definitions of "margin" then there should be two different mathematical representations correspondingly. But as I see it does not matter if "margin" is defined as distance between two support vectors or a separating hyperplane and one support vector, it will have the same formula, namely $2/||w||$. But how is it possible?



Alexandre KOWALCZYK

Post author

March 16, 2016 at 8:49 pm

Hello. The two hyperplanes H_0 and H_1 are not support vectors, they are hyperplanes. As they separate the data, they could be called separating hyperplanes too. In the case where the central hyperplane is the optimal hyperplane, then vectors which lie on the two other hyperplanes will be called support vectors. In the case where the margin is the distance between a support vector and the hyperplane it will have a length of $1/||w||$. It does not matter which one we choose for the margin because at the end we optimize the norm of w and the same argument works if we use $2/||w||$ or $1/||w||$ (minimizing the norm is like maximizing the margin). If you want to understand more about the margin you can take a look at [this lecture](#) of Andrew Ng which introduces the functional and geometric margin. I hope it helps you.



sourena

March 23, 2016 at 3:25 pm

thanks man...really wonderful tutorial...look forward to part 4.



sandeep

March 23, 2016 at 6:59 pm

Such a wonderful explanation any layman can understand SVM. I appreciate your work and the way you explained.



Ma Lu

March 26, 2016 at 1:04 pm

Could you please give a hint about when part 4 could be released, no stress i'm levels just curious 😊



Alexandre KOWALCZYK Post author

April 3, 2016 at 8:12 am

Hello. I honestly don't know. I will try my best to do it in the next few months.



optimus

March 27, 2016 at 7:50 am

Hi Alexandre, could you explain to me what is the correlation between finding hyperplane with prediction process by svm? I have read several articles about svm, but I still don't get it



Alexandre KOWALCZYK Post author

April 3, 2016 at 8:16 am

Hello optimus. This is a two step process: you find any hyperplane which separates data, and you are able to make predictions on your training set. However you have no guarantee that it will perform well in the test set (that it will "generalize" well). With SVMs you find the "optimal" hyperplane, one which separates the data, but which is also as far away as possible from it. As a result it generalizes better and you can expect that the prediction you made on unseen data will be better.

**renjith**

April 1, 2016 at 1:45 pm

brilliant!! eagerly waiting for part 4!!! thanx a ton! 😊

**neran**

April 4, 2016 at 2:14 pm

Really enjoying the article. Thanks for the clear description of something very powerful.

**Daniel**

April 20, 2016 at 7:47 pm

You explained everything so well and in such a non-patronising manner. I wish you were my teacher years ago. Looking forward to part 4.

**Alexandre KOWALCZYK**

Post author

April 26, 2016 at 11:35 pm

Thank you very much Daniel. Part 4 is on its way 😊

**Sourabh**

April 29, 2016 at 4:19 pm

I really liked your post and it gave me a keen interest in SVM, although i do have some open points which I do not understand after reading this 3rd part. I followed till combining of the constraints in one equation(till eq 8). But I have some open points regarding the initial selection of any hyperplane like how we start initially with first hyperplane which will classify the dataset(I know we can do it through single perceptron),but the thing is here in this post when we maximizing the distance in those steps I really gone confused.

thanks for this wonderful post.

It will be great if you can take one very simple 2 dimensional example to explain what we did in part 3 which explain all steps we covered here , it will be very helpful for visualization ,like what is going on.



ross

May 15, 2016 at 1:39 pm

Looking forward to part 4.



Kaustav Mukherjee

May 22, 2016 at 12:57 am

Hey Alexander,

Let me first thank you for an absolutely outstanding explanation of SVM! I have scoured the internet upside down, ranging from lectures from MIT and Caltech profs, none of the explanations I have found can match yours in terms of simplicity and effectiveness! Please upload part 4 ASAP! I am unable to have peace of mind until I see it working end-to-end!



Alexandre KOWALCZYK

Post author

May 24, 2016 at 10:13 pm

Thanks for you kind words. I am working on Part 4. 😊



sawssen

May 30, 2016 at 1:47 pm

Hello, thank you very for this explanation , i need to inderstand the 4 th part , is it ready ?

**Alexandre KOWALCZYK**

Post author

June 6, 2016 at 12:24 pm

Hello. The 4th part is not ready yet but contains already some content, I want to put enough content in it so that people are not frustrated and we the article advance the understanding of SVM. I have some time to work on it this week because I am on holidays s but most of the time I have very little time to write during the week, hence the big delay.

**marcelo**

June 3, 2016 at 4:38 pm

Hello Alexandre,

I have a question about a special case of SVM. I tried to find the support vectors for these training examples:

$$x_1 = (0,1) \quad y_1 = +1$$

$$x_2 = (1,0) \quad y_1 = +1$$

$$x_3 = (-1,0) \quad y_1 = -1$$

$$x_4 = (0,-1) \quad y_1 = -1$$

Clearly, all examples are support vectors because x_1 and x_2 are over the hyperplane H_+ and x_3 and x_4 are over the hyperplane H_- . However, when the LD is maximized subject to $\sum(y_i \cdot \alpha_i = 0)$ and $\alpha_i \geq 0$, the obtained results are (I have used the Wolfram Alpha online optimizer):

$$\alpha_1 = \alpha_3$$

$$\alpha_2 = 1 - \alpha_3$$

$$\alpha_4 = 1 - \alpha_3$$

(notice that α_i is the Lagrange Multiplier of x_i)

The solution that has sense to me is $\alpha_1 = \alpha_3 = 1$, but then $\alpha_2 = \alpha_4 = 0$, indicating that x_2 and x_4 are not support vectors, and this is illogical. Even more with $\alpha_1 = \alpha_3 = 1$ and $\alpha_2 = \alpha_4 = 0$, the w vector and b obtained return these results:

$$(w \cdot x_1) + b = +1 \quad (x_1 \text{ is over the } H_+, \text{ since } x_1 \text{ is a support vector})$$

$$(w \cdot x_2) + b = +1 \quad (x_2 \text{ is over the } H_+, \text{ but it is not a support vector because its Lagrange Multiplier is zero})$$

$$(w \cdot x_3) + b = -1 \quad (x_3 \text{ is over the } H_-, \text{ since } x_3 \text{ is a support vector})$$

$$(w \cdot x_4) + b = -1 \quad (x_4 \text{ is over the } H_-, \text{ but it is not a support vector because its Lagrange Multiplier is zero})$$

I have also used the `svmtrain` of Matlab and the support vectors obtained are the same (x_1 and x_3).

One last observation, if I add one more training example, $[x_5 = (2, 2) \quad y_5 = +1]$, only α_5 is zero, this indicate that x_1 , x_2 , x_3 and x_4 are support vectors.

Do you have any suggestion to clarify this situation?

Thank you in advance.

**Alexandre KOWALCZYK**

Post author

June 12, 2016 at 10:11 pm

Hello marcelo,

I performed the same experiment as you with a quadratic programming library and I have found 4 support vectors. This is because I was solving the problem without regularization (the hard-margin). However most libraries give you an implementation of soft-margin SVM with regularization where the hyperparameter C allows you to tune how much you want to regularize. By default they use a value of C which perform some regularization. I made a reproducible example in Python. If you change the value of C to be greater than one, you will have 2 support vectors. However as soon as you use a value less than 1 you have 4 support vectors. I hope it helps you.

```
1 from sklearn import svm
2 import numpy as np
3
4 X_train = [[0,1],[1,0],[-1,0],[0,-1]]
5 y_train = [1,1,-1,-1]
6
7 # As soon as C becomes greater than 1, the number of SV becomes 2 instead of 4
8 clf = svm.SVC(kernel='linear', C=0.5)
9 clf.fit(X_train,y_train)
10 print('There is {} support vectors'.format(np.sum(clf.n_support_)))
11 print(clf.dual_coef_ )
```

impact_c.py hosted with ❤ by GitHub

[view raw](#)**abood**

June 25, 2016 at 3:27 pm

Hello Alexandre,
thank you for an absolutely outstanding explanation of SVM
have you Code in Java for SVM !!!???

**Alexandre KOWALCZYK**

Post author

June 25, 2016 at 7:39 pm

Thanks abood. There is a [java implementation](#) of libsvm.

**abood**

June 26, 2016 at 2:57 pm

Thank you very much Alexandre for the java Code.
But I want to java Code only for linearly separable SVM please

**Chew**

June 6, 2016 at 4:56 pm

Very, very good explanation. Great Job!!!!

You have a book about this? If yes send-us the name, I wanna buy it!

Hope you can explain us the next steeps soon! I am excited with your step-by-step, figures and numerical examples.

**Alexandre KOWALCZYK**

Post author

June 7, 2016 at 3:19 pm

Thanks a lot for your comment. I don't have a book yet 🙄 The next part is coming up. Don't forget to drop your email and I will keep you informed.

**G. Krishna veni**

June 10, 2016 at 10:26 am

Sir really i satisfied with u r modules. Actually I did not have basic maths. When I gone through this blog i cleared doubts myself. I'm eagerly waiting for your next part

**ntp**

July 4, 2016 at 9:29 am

If you write a book about Machine Learning with these kind of awesome explanation, I will buy it. Anyone has a same thought?

**abood**

July 6, 2016 at 6:27 pm

Helle Alexandre,
I want java Code for linearly separable SVM please!!!

**neoli365**

July 12, 2016 at 4:19 am

Thank you very much for this clear, awesome explanation. Looking for part 4 ~~

**Cam**

July 17, 2016 at 3:16 pm

Excellent tutorial. I'm looking forward to part 4.

Cam

**SUMAN KUMAR CHOUDHURY**

July 24, 2016 at 8:58 am

Great !!! wao !!! what an explanation... thank u very much.... I am waiting for the optimization in SVM (Part 4)



Jorge

August 15, 2016 at 6:33 am

Thank you very much! It's great to read a thorough description of the SVM with a clear notation and the intuition behind the math.



Arun BM

August 16, 2016 at 11:13 am

This was best explanation ever....Now I need not see any vedios books Thanks a Lot



Rimple

August 21, 2016 at 8:39 pm

Excellent tutorial. Best explanation of SVMs. Eagerly waiting for part 4



Cor

September 12, 2016 at 2:37 am

Hi, thanks for writing this up. Could you please double-check that w is indeed perpendicular to H_1 ? If you take any point on H_1 , this point is represented by a vector drawn from the origin to that point. This vector is obviously not perpendicular to w . I'd appreciate it if you could explain why I'm wrong 😊



Alexandre KOWALCZYK

Post author

September 12, 2016 at 8:27 am

The two things that matter about w are its direction and its magnitude. I can draw it where I want and I will be the same vector (even if it does not have the same starting coordinates). I explain this in Part 2 when talking about the difference between two vectors. I hope it helps you 😊



Bahareh Moradi

September 28, 2016 at 10:39 am

Hi!

In equation 9 we have $K = \mu = m (w / ||w||)$;

I think in defining the vector " k " we need something like b in $wx+b=0$

to indicate the location of vector in the space.

just the magnitude and the norm shows where is the starting point of k ????

Please help me



Alexandre KOWALCZYK

Post author

September 28, 2016 at 10:43 am

Hello. No, you do not need to indicate the location of the vector in the space. This is often confusing, I explain this in Part 2 when talking about the difference between two vectors. Reread this section, it might help you understand.



Amal Targhi

October 12, 2016 at 1:42 pm

Thank's for this course but i can't understand how do you calculate the vector W and the value of b in order to know if $wx+b$ superior or inferior to 0 please understand me



Alexandre KOWALCZYK

Post author

October 13, 2016 at 11:20 pm

The value of w is what we are trying to find by solving the optimization problem explained in the following articles.



Sonu Lamba

November 3, 2016 at 7:49 pm

Dear Sir,

How we calculate w and b ? are they different for every feature and at the end we need to optimize them?..how we decide the the value of w and b for optimal hyperplane equation...if for each feature we are calculating w and b then what value of w and b we will choose?. and why we take unit value for w always even the data are not always at the same distance. I am not getting the base of w and b value. could you please explain it earliest. Thank you very much for your time and article.



Alexandre KOWALCZYK

Post author

November 3, 2016 at 9:23 pm

Hello. There is only one vector w , and it will contain several values. If X is an (m,n) matrix, then a single example is a $(1,n)$ vector and w is a $(n,1)$ vector. So one vector w contains the values for all examples. There is a single value of b . They will be chosen by solving an optimization problem (you can read more on that on the later articles).



cam nguyen

November 5, 2016 at 6:54 pm

Hello Alexandre, thank you so much for your great article !

I'm learning a project about human detection which using SVM algorithm to classify non-human or human. Can you tell me the relationship between the variables w , b , x_i with the values from the input images? I haven't still imagined yet? TKS so much!

Alexandre KOWALCZYK

Post author



November 6, 2016 at 3:25 pm

Hello. This is a pretty broad question and the answer depends a lot on how you transform your image. I advise you to read this [paper](#) to learn more about image classification with SVM.



cam nguyen

November 9, 2016 at 4:14 pm

How to compute the weight vector w and bias b in linear SVM. We have a hyperplane equation and the positive and negative feature. In equation $Wx+b=0$, what does it mean by weight vector and how to compute it?? Can anybody explain it please



Matthias

December 1, 2016 at 10:58 am

Hello. I agree with the many comments praising your post. Thank you so much for it!

There's one thing I'm still confused about: How do we find a hyperplane that separates the set in the first place? In step 2 it says: "Given a hyperplane H_0 separating the dataset and satisfying[...]" So how do we get H_0 ? Did I somehow miss that?



Alexandre KOWALCZYK

Post author

December 4, 2016 at 6:43 pm

Hello, Matthias. Thank you for your comment. The hyperplane H_0 is introduced in order to derive the optimization problem. I could start an argument by "Given a point x on the horizontal axis", and it does not matter how you pick it, or where it is, what matters is that it is a point, and it is on the horizontal axis. The same apply in this context. But to answer your question, if you have a linearly separable dataset, you can find a separating hyperplane by using an algorithm like the Perceptron Learning Algorithm for instance, or of course the SMO algorithm for the SVM problem.

**Matthias**

December 12, 2016 at 1:15 pm

Thank you so much for clarifying. That was very helpful!

**Kamakshi Bansal**

December 20, 2016 at 7:33 pm

You are such an amazing teacher!! kudos to this post. This post taught me SVM in such a simple way.

**trung**

January 7, 2017 at 8:04 pm

thank you so much, you saved my whole day researching how the distance is computed and now I found it.
thank you again!

**Vijay Gupta**

January 13, 2017 at 6:06 am

Amazing !!!!

**idk222**

January 24, 2017 at 4:32 pm

What is the intuition for why minimizing the norm of w is the same as maximizing the margin m . From the algebra it is obvious.

w is the normal vector to H_1 . The norm of w is the magnitude of that vector. Do you have any intuition why that would be the same as choosing the biggest m ? Thanks. Great articles!!

**Alexandre KOWALCZYK**

Post author

January 26, 2017 at 8:10 pm

I already thought about this but I did not come with a simple intuition. I would be glad to hear one if someone has a good explanation.

**Numhacker**

February 1, 2017 at 2:25 am

It's all about dot product. Normal vector w is perpendicular to the hyperplane, and so the margin vector k is. Given arbitrary vector w and k , what we want is the result of this dot product must be 1 (or any positif number, we use δ in this post). Since these two vectors is in parallel, the dot product is only based on its length. How do you choose w and m which the dot product is 1 while we want to maximize margin? Yap, minimize the length of the normal vector w !

**Tetsuya**

February 4, 2017 at 3:47 am

Thanks a lot for the detailed explanation. That's the greatest article on SVM I have ever read before. Here I have one question. In Figure 4, 5, 6, 7, 8, there are hyperplanes written by $wx - b = 1$ and $wx - b = -1$, but we are talking about $wx + b = 1$ and $wx + b = -1$, right? So, I think that $wx + b = 1$ and $wx + b = -1$ should be also written in those figures. What do you think?

**Alexandre KOWALCZYK**

Post author

February 15, 2017 at 11:05 pm

Yes indeed it would be more coherent to use $wx+b=1$ in the figure as well. Even if the only difference between an hyperplane with $wx-b=1$ and $wx+b=1$ is the sign of b . Thanks for pointing that out. It is in fact often confusing in the literature as well as some author use the $wx+b$ convention and others use $wx-b$ (Platt for example).



ridarafisyed

February 19, 2017 at 11:34 am

the w used here is also direction vectors?



Alexandre KOWALCZYK

Post author

February 23, 2017 at 9:40 pm

I am not sure to understand your question perfectly. But the vector w here is the vector normal to the hyperplane. In Part 2 we defined a vector w which was the direction of a vector u . But they are not the same vectors.



Sam

February 28, 2017 at 3:49 am

Can I upvote the article? I mean, I would click on the upvote icon til my mouse blows off.



Rahul

April 4, 2017 at 2:51 pm

Simply Amazing. Sorted out all the doubts regarding SVM. Can you please take any classification example along with say MATLAB code without using any direct command?

Thanks

**Alexandre KOWALCZYK**

Post author

April 4, 2017 at 7:42 pm

Thank you Rahul. Unfortunately my MATLAB is a little rusty 😊

**Nick**

April 20, 2017 at 7:36 pm

You wrote in a previous tutorial: "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"

However with the constraints above(in this tutorial) there isn't guarantee that a data point (of the training set) was on the boundary of hyperplane H_0 or H_1 . In this case the given definition of margin seems wrong. Am I wrong?

Another question is why $\delta=1$? It should be correlation between the margin and δ . If you leave δ you get (if I'm not wrong):

$$\text{margin} = (2 \cdot \delta) / ||w||$$

And the optimization problem should involve the parameter δ too.

Why can simplification be done ($\delta=1$)?

**Alexandre KOWALCZYK**

Post author

April 24, 2017 at 10:17 pm

Hello Nick,

1) As the hyperplane will be at equal distance between the two nearest data point finding the distance from one point than multiplying it by 2 works.

However this is not very rigorous, that is why we see the "true" method to compute the margin in this article.

2) To better understand delta, you can read [this paper](#). (Equation 2)

**Tushar**

April 27, 2017 at 6:22 pm

Great work. I am a Deep Learning researcher. I personally have never seen such simple and effective explanation for SVMs in any of the books (or across the web). This reminded me of the quote by Einstein...."If you cannot explain it simply, you probably have not understood it"Cheers.

**Mariusz**

May 1, 2017 at 12:47 pm

Hi there. I really liked this article! As a reference for the future I created this simple visualization. Maybe somebody else will find it useful:

<https://www.desmos.com/calculator/4jqnpapcv9>

**Alexandre KOWALCZYK**

Post author

May 14, 2017 at 6:14 pm

Thanks a lot Mariusz!

**Joey**

May 2, 2017 at 5:25 am

great! thank you.

**Olle Welin**

May 11, 2017 at 5:59 pm

Hi thank's very much for your tutorial. Very informative. (As usual when I got into new intellectual topics I often miss understand basics). Regarding SVM I wonder how will I do to find the first 3 support vector. One guess I have is this (correct me if I am wrong).

1. Create a first version of test "nr1 hyperplane" by Random select an origo and a random angle for the (bolded) '**w**' vector (hyperplane)
2. Go through All vectors and if some vector is constrained by this test "nr1 hyperplane" then throw this vector.

Maybe see how bad this constrained point and use this data to produce a new test "nr2 hyperplane".

3. When test "nr? hyperplane" satisfy ALL vectors then Start to find the 3 support vectors.

4. Now again go through the margin calculation method you describe for ALL the vectors and when some vector have smallest margin save this until you have find for example 'A' 'B' and 'H' in your example above.

I guess there must be a way to find this 3 support vector as fast as possible.

Very best Regards Olle Welin

I feel so slow but I am curious about machine learning how the basics math "nuts and bolts" works. (play with write my own C code on Raspberry pi)



Alexandre KOWALCZYK Post author

May 14, 2017 at 6:26 pm

Hello. If you want to solve the problem in a fast way, you should read [Platt's SMO algorithm](#) paper 🤔



Mohit

May 21, 2017 at 9:37 am

Great work Alexandre! Sorry m completely new to this so might be a novice question but am not able to visualise how z_0 which is addition of x_0 and k is in hyperplane H_1 ?

"If we start from the point x_0 and add k we find that the point $z_0 = x_0 + k$ is in the hyperplane H_1 ?"



Alexandre KOWALCZYK Post author

May 22, 2017 at 7:29 pm

Well, you do not need to visualize it because you can **see** it on Figure 14. There is a review of vector addition on part 2 if you need help understanding the concept. Best Regards,



Hannah Liu

May 25, 2017 at 6:54 am

I am still confused about step 2.

Yes, with all the constraints, I will be able to know whether the hyperplanes are the ones that correctly separate the data. But how do I get the hyperplanes in the first place? Do I take a random guess? If after applying the constraints I found they are not the ones I am looking for, how do I improve myself?

**Alexandre KOWALCZYK**

Post author

May 27, 2017 at 7:20 pm

That is the purpose of the [algorithm SMO](#). Right now this article is building the foundation used to define the problem we want to solve, not explaining how to solve it. (I will explain all of that in my upcoming ebook)

**Peyman Barjoeian**

June 19, 2017 at 1:56 am

Hello Dear Alexandre.

Thank you for these good series.

I have a question. it make sense to minimize $||w||$ value in our optimization problem. but why should also minimize the b value? how they($||w||$ and b) are related?

Cheers.

**Alexandre KOWALCZYK**

Post author

July 23, 2017 at 9:15 pm

I think [this page](#) might help you understand the importance of b.

**Dewei**

September 28, 2017 at 11:42 pm

Thanks for the detailed explanation. It's super helpful!



JX Song

January 3, 2018 at 9:02 am

First of all I'd like to thank you for this awesome tutorial. There is however something I do not understand. In the part "Understanding the constraints". Why is it that for $\xi = A$, the point on the hyperplane is $=1$ and for $\xi = C$ the point on the hyperplane is >1 ?



Miruna Stefan

January 24, 2018 at 10:30 pm

Dear Alexandre,

I have been circling SVM for over a year now, and while i understood the underlying principle, i never truly got how it worked. Most resources provide either a general, high-level explanation, or heavy mathematics that can be difficult to understand on your own.

Your post, on the other hand, is clear, detailed and explicit. I think you have invested a great deal of attention and time into this project, and i want you to know it is truly helpful and of great quality.

Thank you very much!

Miruna



Romain C

January 29, 2018 at 2:46 pm

Hi,

Thanks again for this great tutorial,

I got to the end of the first tutorial steps and I am trying to go further reading your ebook on SVMs

I do not understand why :

- here we minimize $\|w\|$
- on your ebook we minimize $1/2 \|w\|^2$

Thanks in advance for your help



Alexandre KOWALCZYK Post author

January 29, 2018 at 9:50 pm

Hello Romain,

I explain this in the book "The factor has been added for later convenience when we will use QP solver to solve the problem and squaring the norm has the advantage of removing the square root."

Best regards,



Steve Chamales

February 19, 2018 at 8:50 pm

I've read the first 3 parts to this tutorial (this page is part 3). In the example you're working through, you chose your first hyperplane to be $x_2 = -2 * x_1$ because this is a very basic example that you could visualize. How do you select a hyperplane that best separates multi-dimensional data that can't be visualized by humans?



Alexandre KOWALCZYK Post author

February 24, 2018 at 1:10 am

Basically we can say that the first hyperplane is defined by a vector w starting at 0 or with random values, then an algorithm updates it by tiny increment until it find the optimal hyperplane (you can read the full detail of this in the chapter about the SMO algorithm in my ebook).

**Yury Parfenov**

March 2, 2018 at 10:10 am

Hi,

We have two parameters w and b in $w \cdot x + b = -1$, $w \cdot x + b = 1$ equations. w is corresponding for rotation, b is corresponding for offset. I didn't get how we can control distance between two hyperplanes if we have shared b parameter and constant values on the right hand side $(-1, 1)$. We only can rotate and shift hyperplanes by varying w and b . But we cannot change distance between planes.

**Alexandre KOWALCZYK**

Post author

March 3, 2018 at 2:12 pm

Compare the distance between the lines created by:

First equation $9.0x + 1.0$ Second equation $9.0x + 3.0$

and by

First equation $0.1x + 1.0$ Second equation $0.1x + 3.0$

Clearly by only shifting the hyperplane we can make the distance smaller.

**Tejveer Singh**

March 24, 2018 at 8:51 pm

Please, can you explain step 1 in a bit detail?

Please explain equation of Hyperplane in step -2 also?

It is a bit unclear in the previous post also.

**dr. hala omar**

May 6, 2018 at 1:06 pm

Thank you very much

your simple and clear explanation helped me a lot to understand the subject especially as a beginner in it.
