

# 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*.

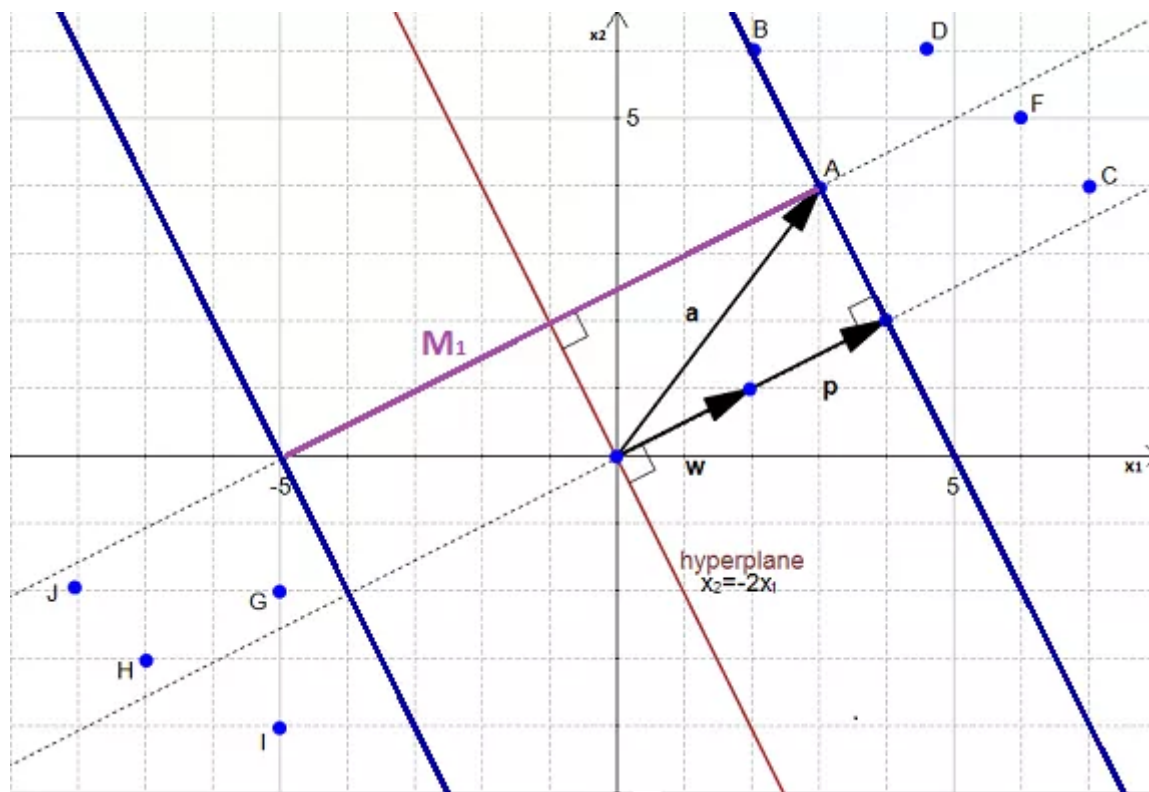


Figure 1: The margin we calculated in Part 2 is shown as  $M_1$

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.

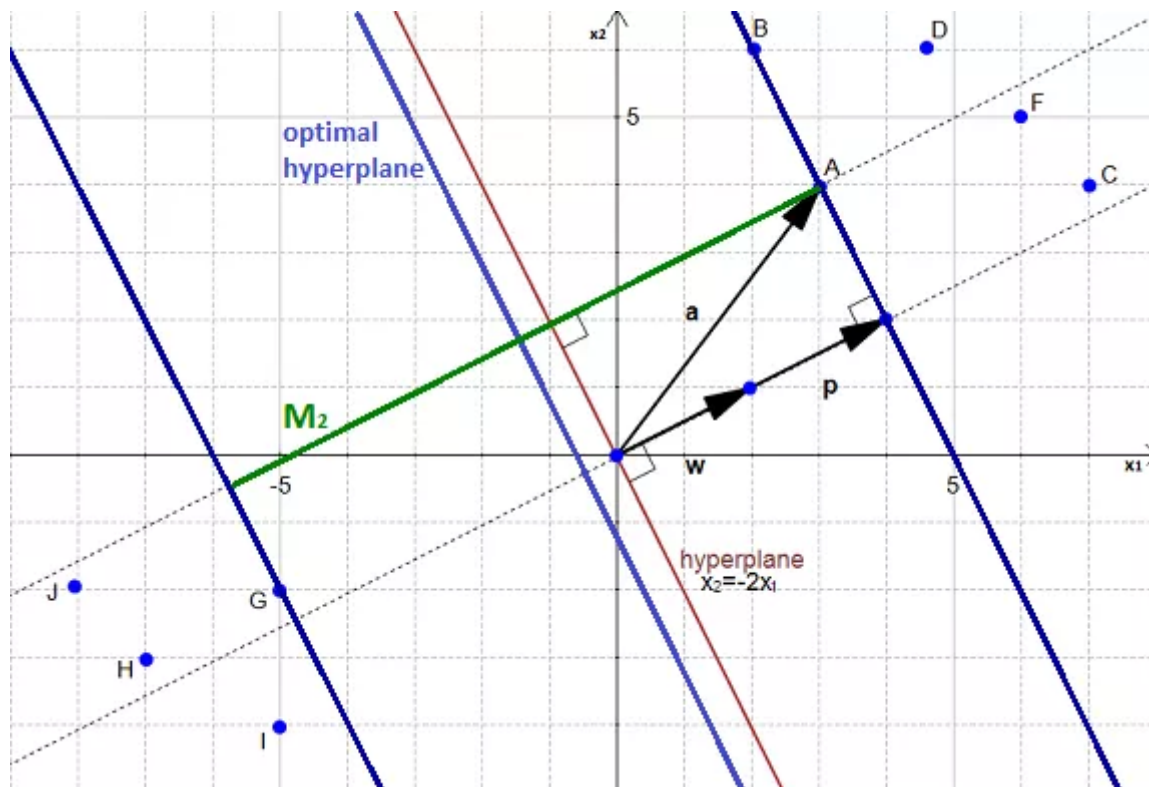


Figure 2: The optimal hyperplane is slightly on the left of the one we used in Part 2.

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*

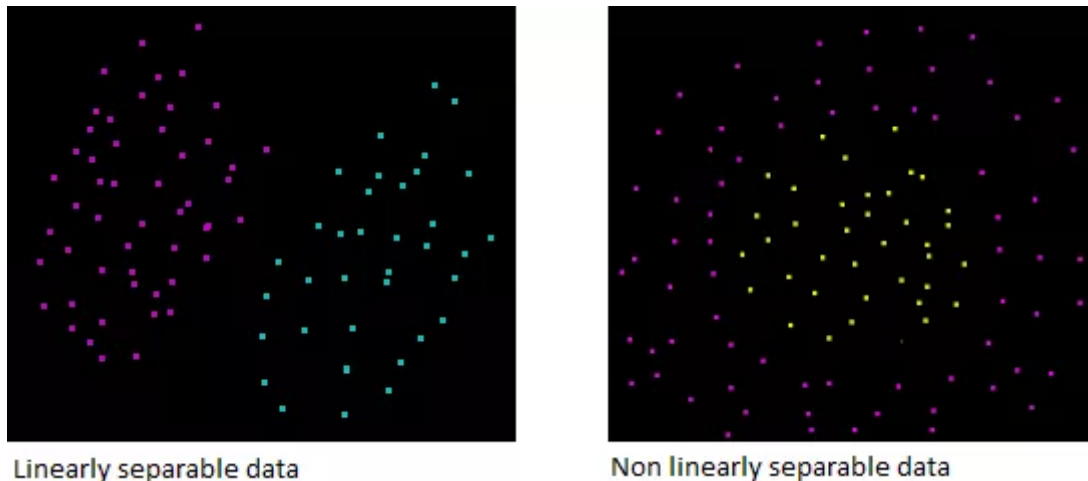


Figure 3: Data on the left can be separated by an hyperplane, while data on the right can't

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  $\delta$  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  $-1$ .

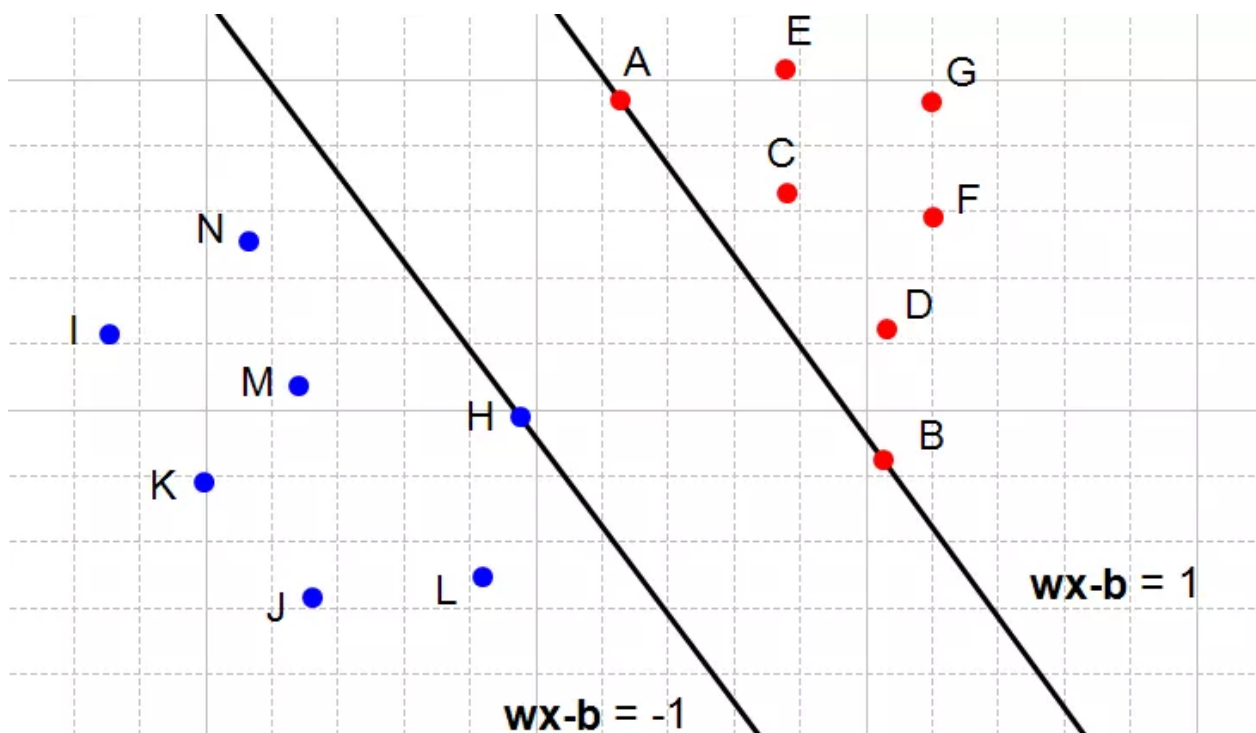


Figure 4: Two hyperplanes satisfying the constraints

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

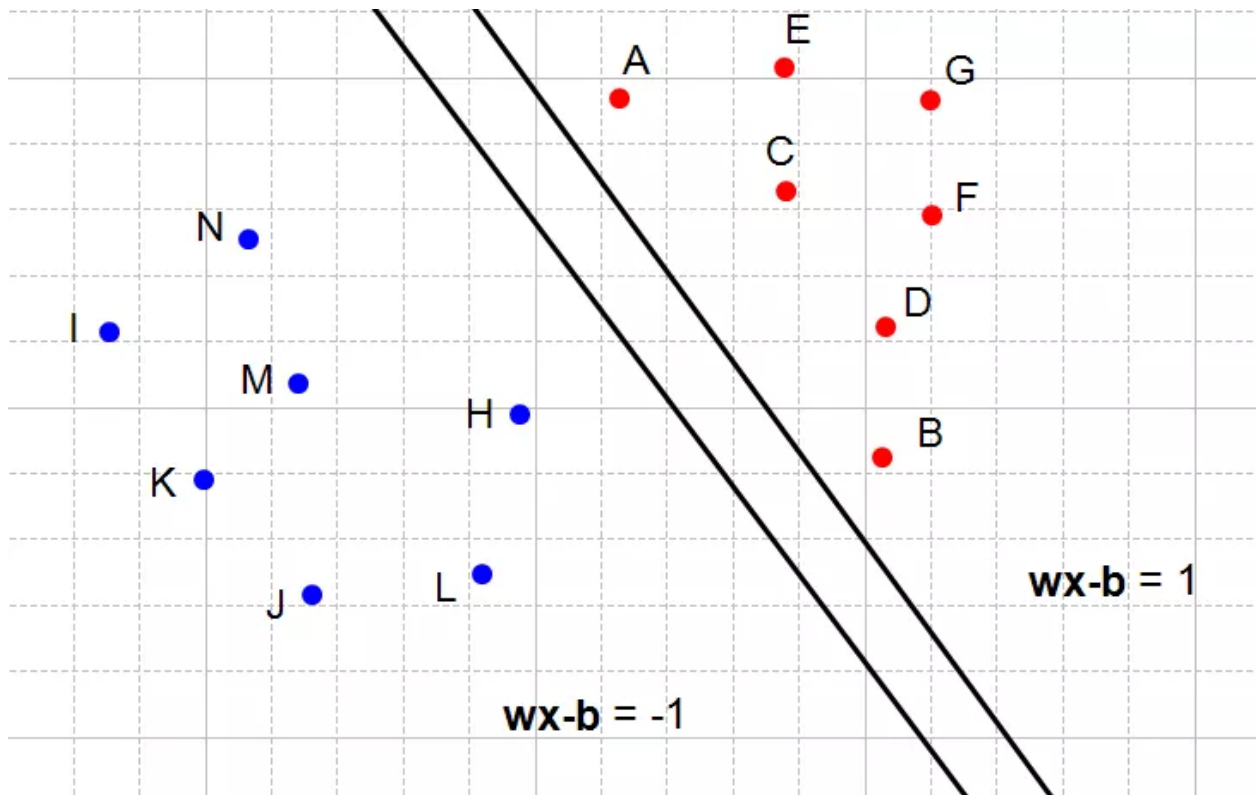


Figure 5: Two hyperplanes also satisfying the constraints

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

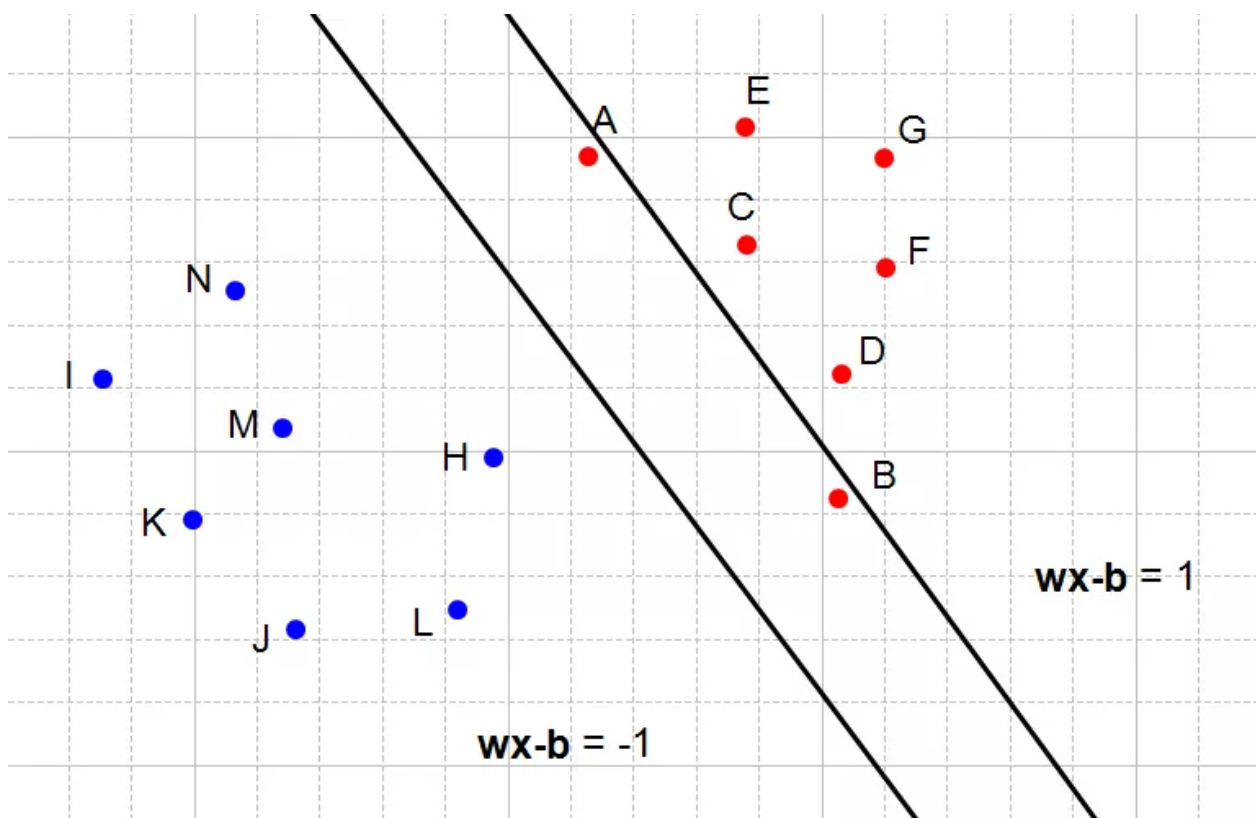


Figure 6: The right hyperplane does not satisfy the first constraint



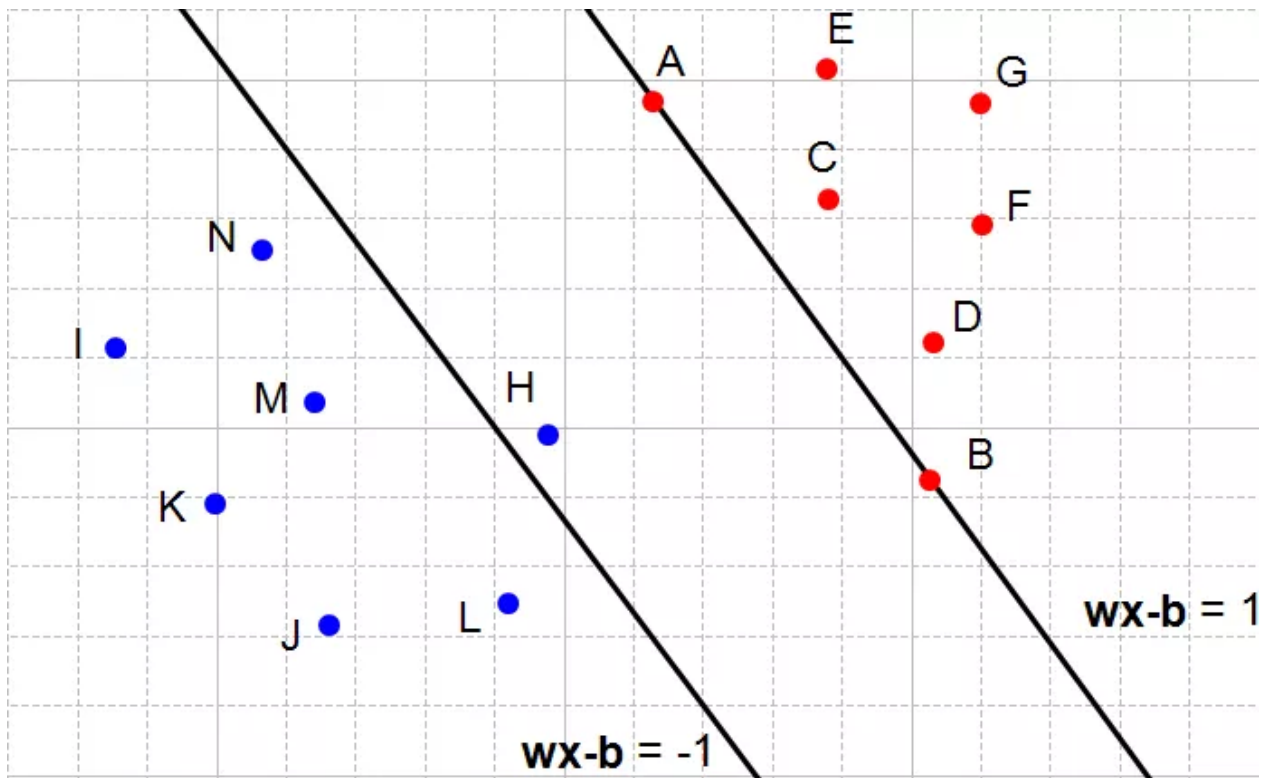


Figure 7: The left hyperplane does not satisfy the second constraint

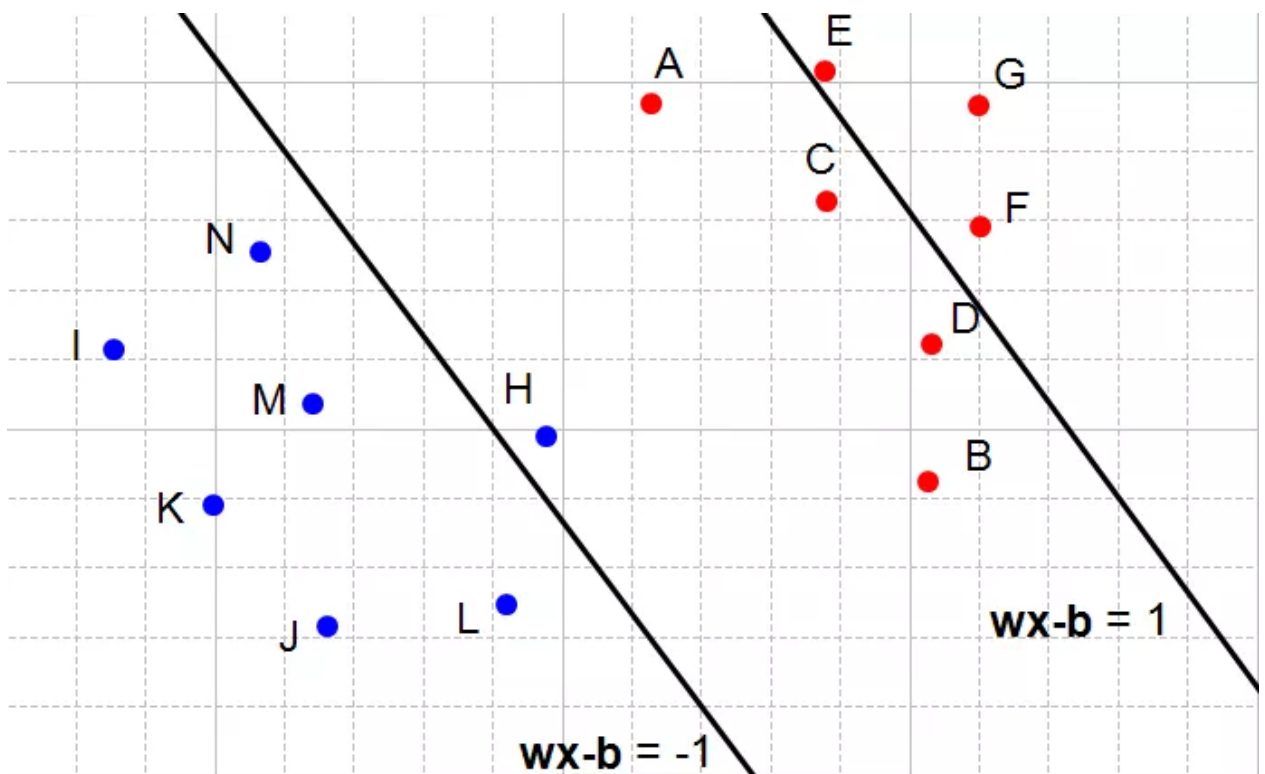


Figure 8: Both constraint are not satisfied

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  $m$ .**

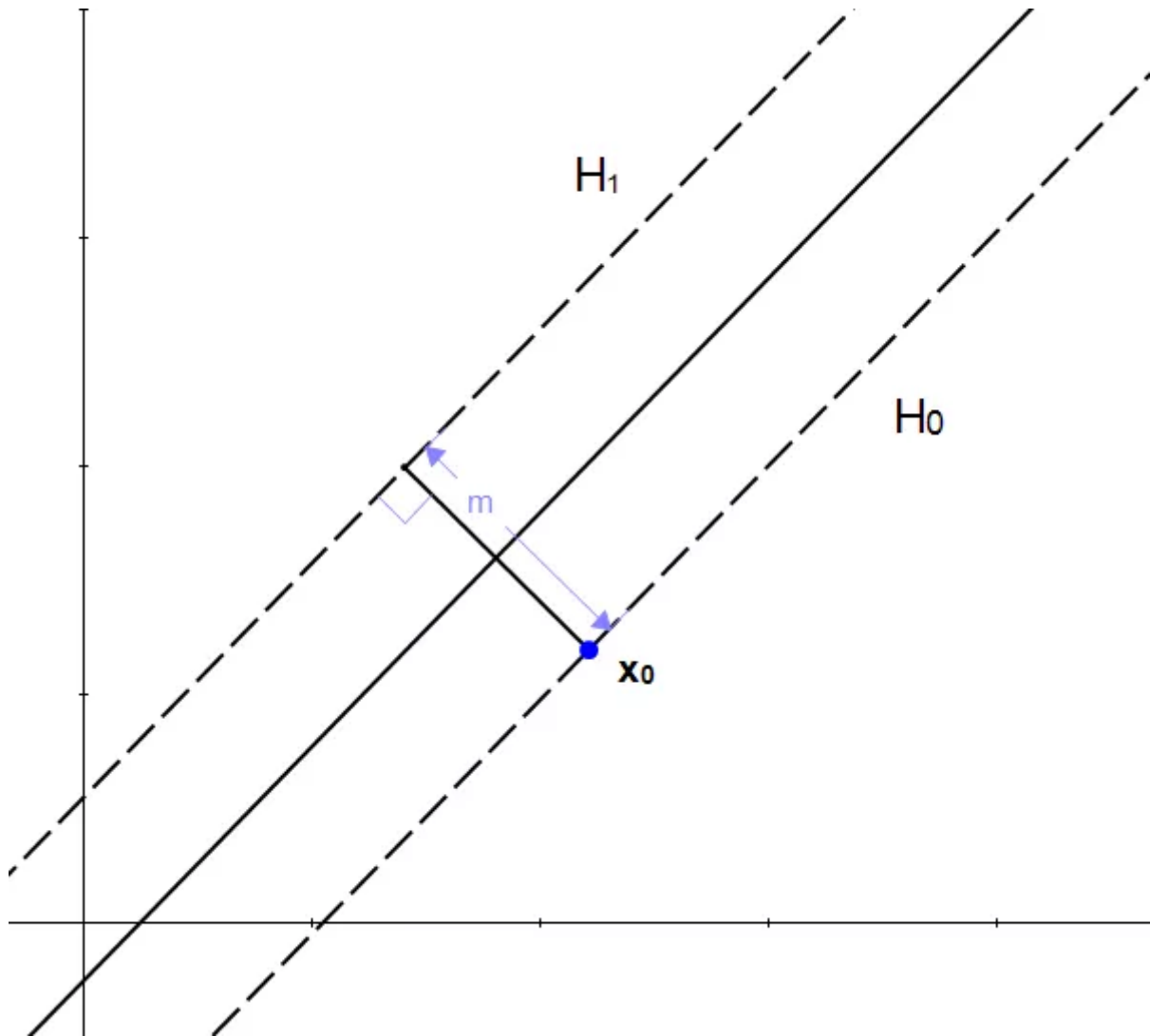


Figure 9:  $m$  is the distance between the two hyperplanes

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 :

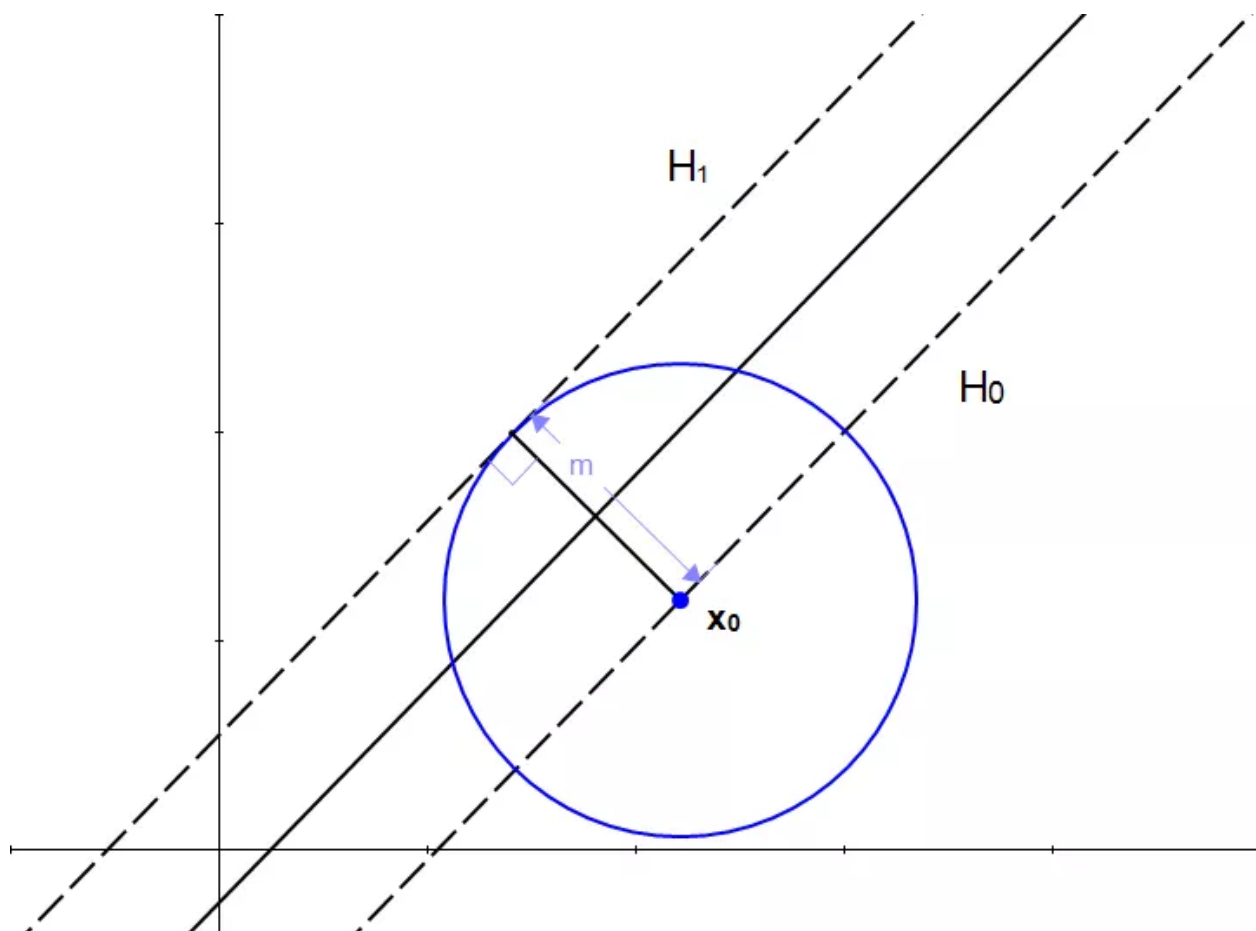


Figure 10: All points on the circle are at the distance  $m$  from  $x_0$

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  $\mathcal{H}_1 = \mathbf{w} \cdot \mathbf{x} + b = 1$  )

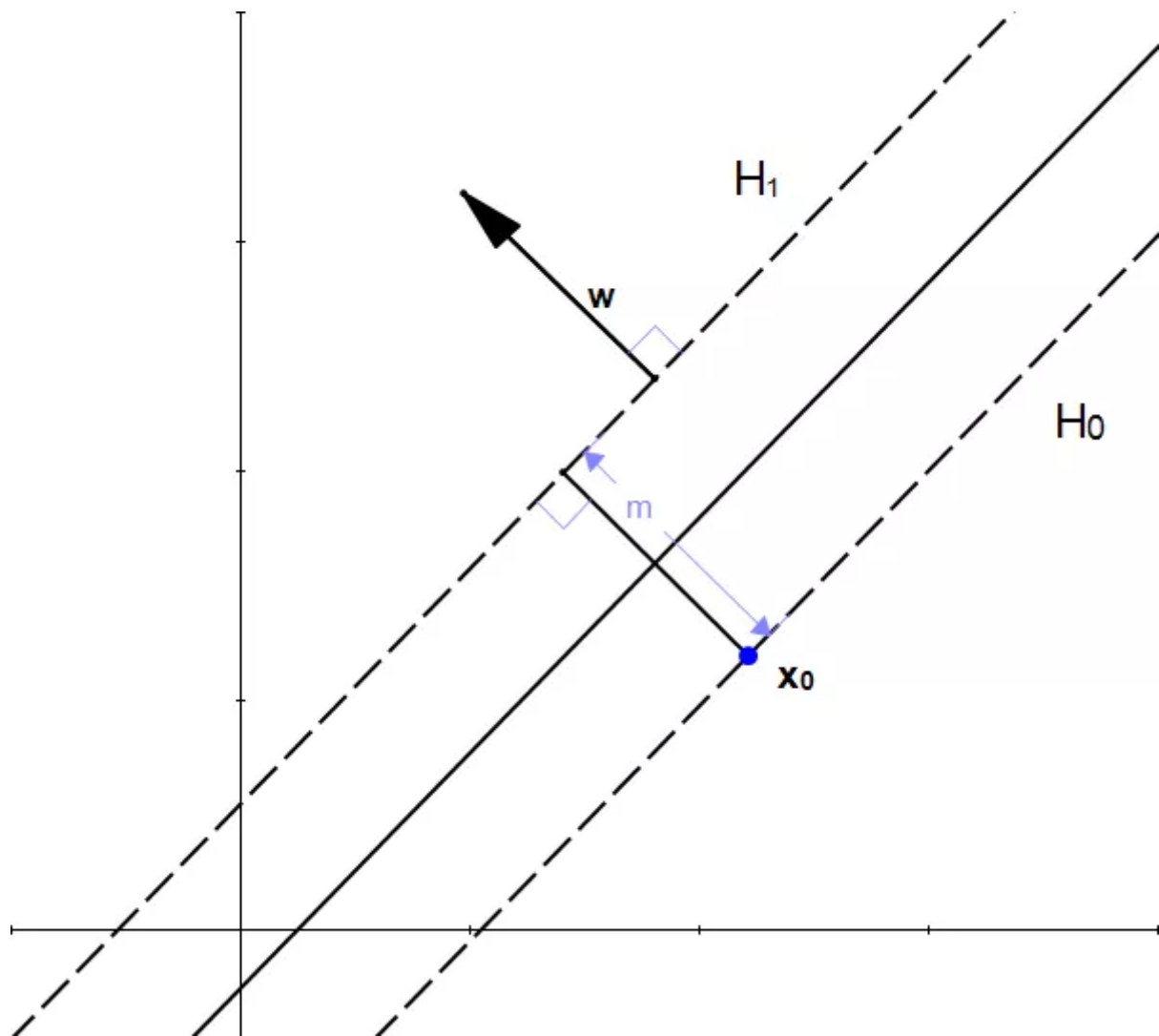


Figure 11:  $w$  is perpendicular to  $H_1$

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.

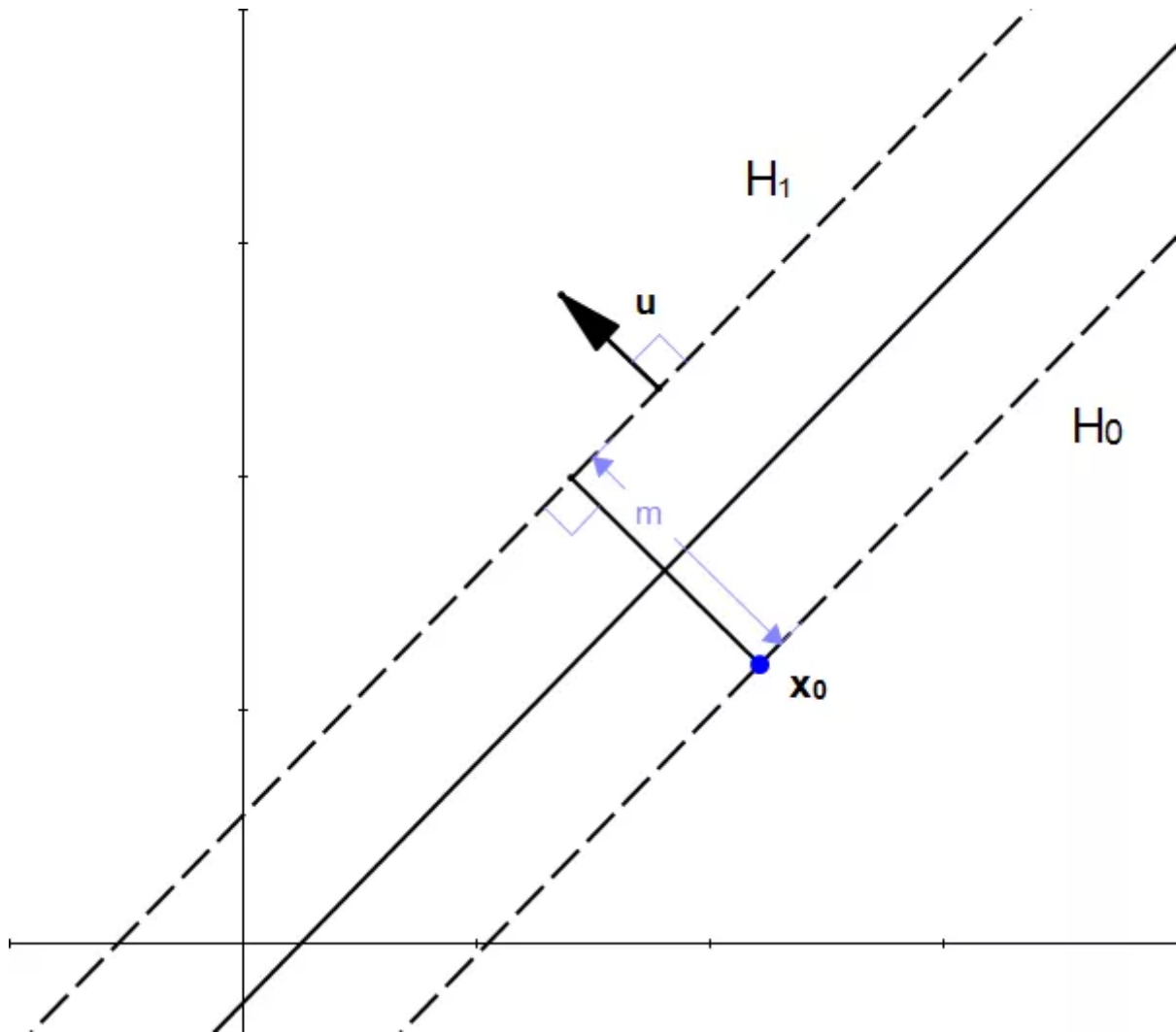


Figure 12:  $u$  is also is perpendicular to  $H_1$

If we multiply  $u$  by  $m$  we get the vector  $k = mu$  and :

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

From these properties we can see that  $k$  is the vector we were looking for.

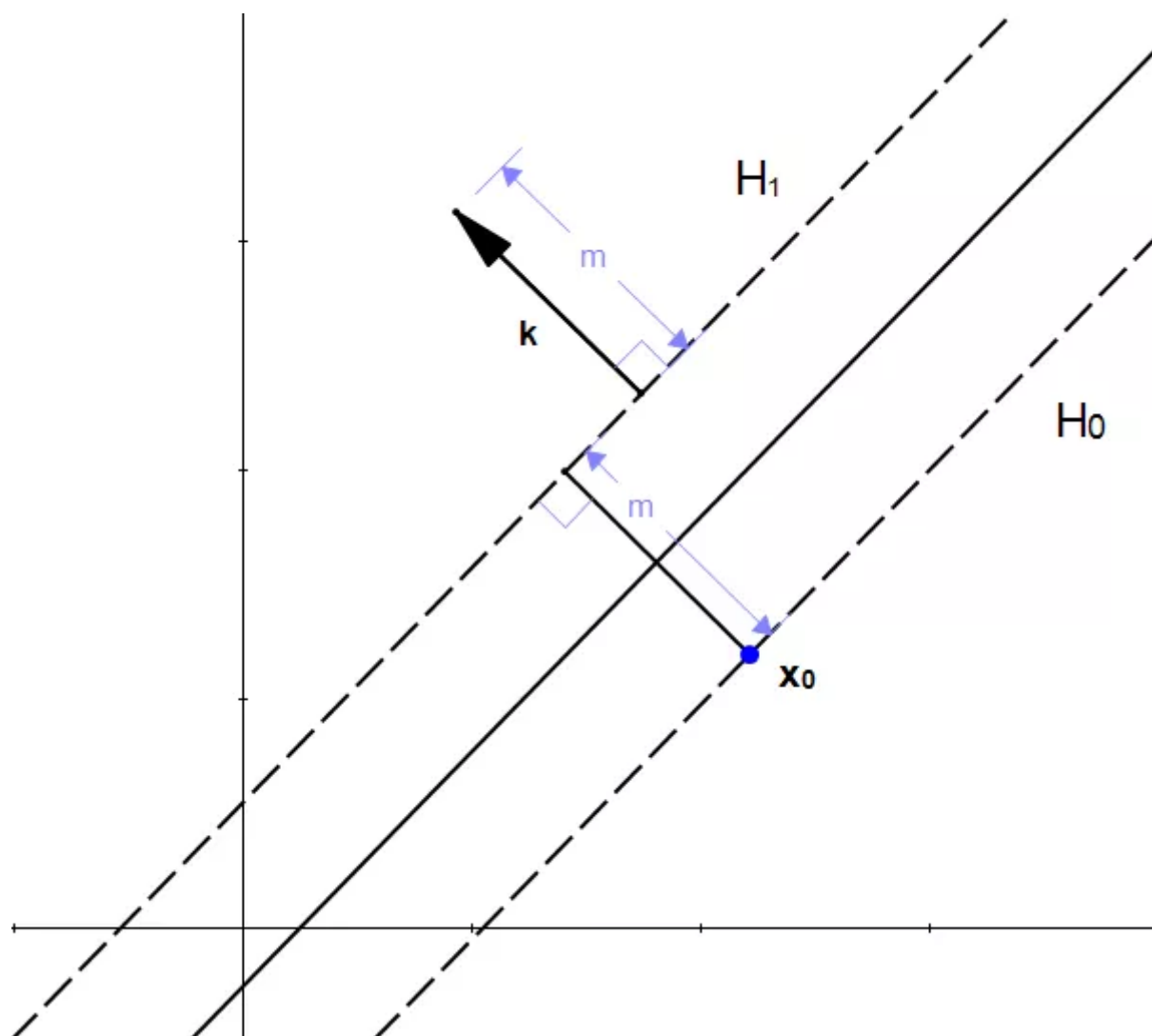


Figure 13:  $k$  is a vector of length  $m$  perpendicular to  $H_1$

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

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  $\mathbf{k}$  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.

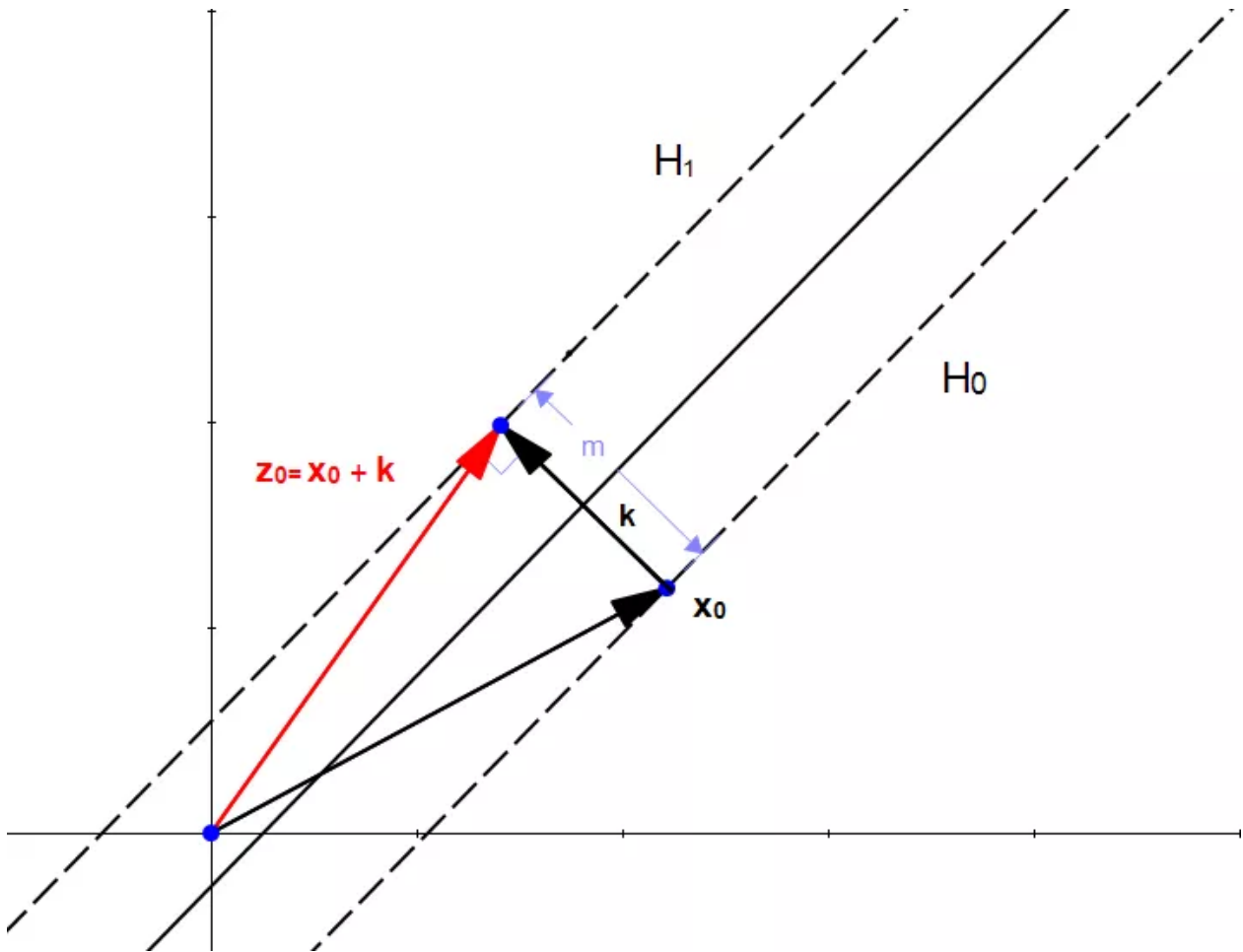


Figure 14:  $z_0$  is a point on  $H_1$

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](#):

$$\text{Minimize in } (\mathbf{w}, b)$$

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

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

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

Solving this problem is like solving an 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/>].

## 149 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