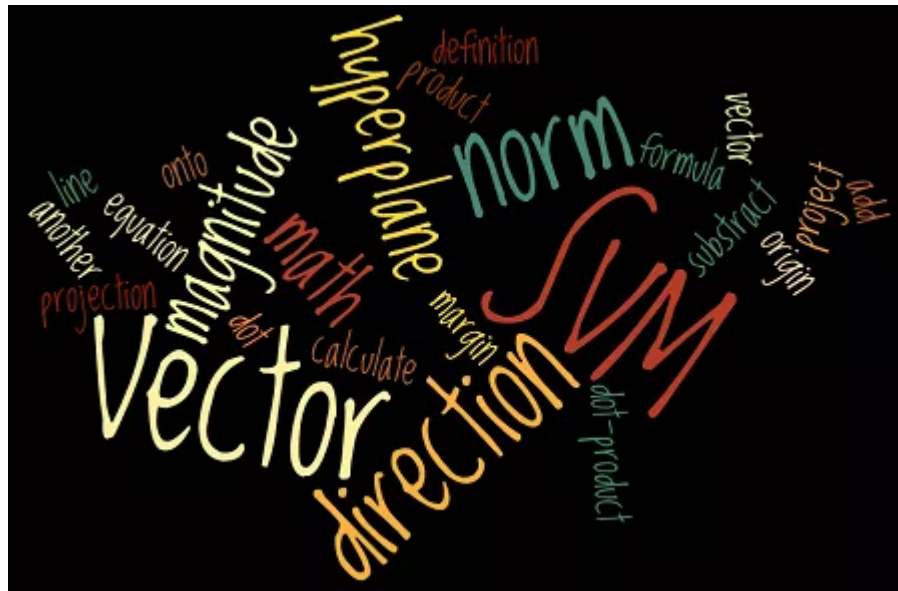# SVM Tutorial

SVM - Understanding the math - Part 2

This is **Part 2** of my series of tutorial about the math behind Support Vector Machines.
If you did not read the previous article, you might want to start the serie at the beginning by reading this article:
an overview of Support Vector Machine.

In the first part, we saw what is the aim of the SVM. Its goal is to find the hyperplane which maximizes the margin.

But how do we calculate this margin?

SVM = Support VECTOR Machine

In Support Vector Machine, there is the word **vector.**
That means it is important to understand vector well and how to use them.

Here a short sum-up of what we will see today:

- What is a vector?
    - its norm

- its direction
- How to add and subtract vectors ?
- What is the dot product ?
- How to project a vector onto another ?

Once we have all these tools in our toolbox, we will then see:

- What is the equation of the hyperplane?
- How to compute the margin?

What is a vector?

If we define a point $A(3, 4)$ in $\mathbb{R}^2$ we can plot it like this.
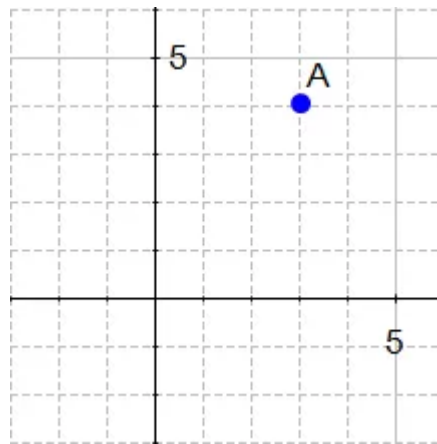

Figure 1: a point

Definition: Any point $x = (x_1, x_2), x \neq 0$, in $\mathbb{R}^2$ specifies a vector in the plane, namely the vector starting at the origin and ending at x.

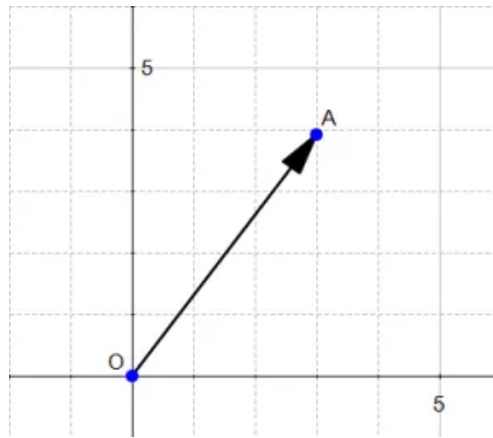This definition means that there exists a vector between the origin and A.

Figure 2 - a vector

If we say that the point at the origin is the point $O(0,0)$ then the vector above is the vector $\overrightarrow{OA}$. We could also give it an arbitrary name such as **u**.

Note: You can notice that we write vector either with an arrow on top of them, or in bold, in the rest of this text I will use the arrow when there is two letters like $\overrightarrow{OA}$ and the bold notation otherwise.

Ok so now we know that there is a vector, but we still don't know what **IS** a vector.

> *Definition: A vector is an object that has both a magnitude and a direction.*

We will now look at these two concepts.

1) The magnitude

The magnitude or length of a vector x is written $\|x\|$ and is called its norm.
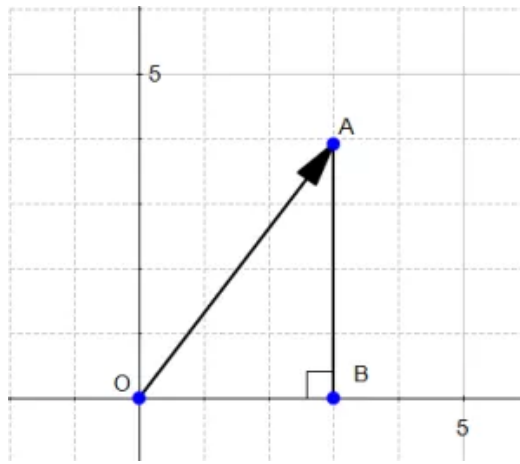For our vector $\overrightarrow{OA}$, $\|OA\|$ is the length of the segment OA

Figure 3

From Figure 3 we can easily calculate the distance OA using Pythagoras' theorem:

$$OA^2 = OB^2 + AB^2$$

$$OA^2 = 3^2 + 4^2$$

$$OA^2 = 25$$

$$OA = \sqrt{25}$$

$$\|OA\| = OA = 5$$

2) The direction

The direction is the second component of a vector.

*Definition : The **direction** of a vector $\mathbf{u}(u_1, u_2)$ is the vector $\mathbf{w}\left(\frac{u_1}{\|u\|}, \frac{u_2}{\|u\|}\right)$*

Where does the coordinates of $\mathbf{w}$ come from ?

Understanding the definition

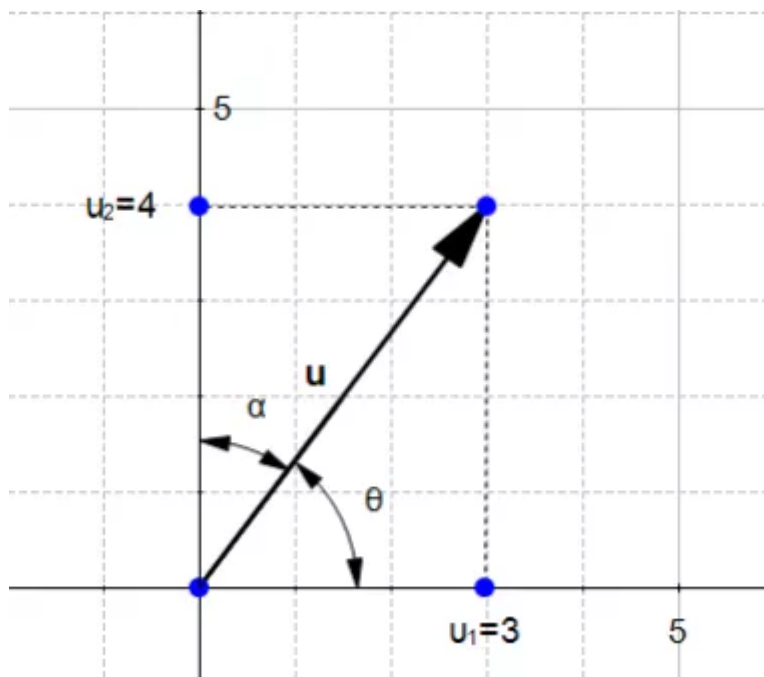To find the direction of a vector, we need to use its angles.

Figure 4 - direction of a vector

Figure 4 displays the vector $\mathbf{u}(u_1, u_2)$ with $u_1 = 3$ and $u_2 = 4$

We could say that :

Naive definition 1: The direction of the vector $\mathbf{u}$ is defined by the angle θ with respect to the horizontal axis, and with the angle α with respect to the vertical axis.

This is tedious. Instead of that we will use the cosine of the angles.

In a right triangle, the cosine of an angle β is defined by :

$$cos(\beta) = \frac{adjacent}{hypotenuse}$$

In Figure 4 we can see that we can form two right triangles, and in both case the adjacent side will be on one of the axis. Which means that the definition of the cosine implicitly contains the axis related to an angle. We can rephrase our naïve definition to :

Naive definition 2: The direction of the vector $\mathbf{u}$ is defined by the cosine of the angle θ and the cosine of the angle α.

Now if we look at their values :

$$cos(\theta) = \frac{u_1}{\|u\|}$$

$$cos(\alpha) = \frac{u_2}{\|u\|}$$

Hence the original definition of the vector **w**. That's why its coordinates are also called *direction cosine*.

Computing the direction vector

We will now compute the direction of the vector **u** from Figure 4.:

$$cos(\theta) = \frac{u_1}{\|u\|} = \frac{3}{5} = 0.6$$

and

$$cos(\alpha) = \frac{u_2}{\|u\|} = \frac{4}{5} = 0.8$$

The direction of $\mathbf{u}(3, 4)$ is the vector $\mathbf{w}(0.6, 0.8)$

If we draw this vector we get Figure 5:

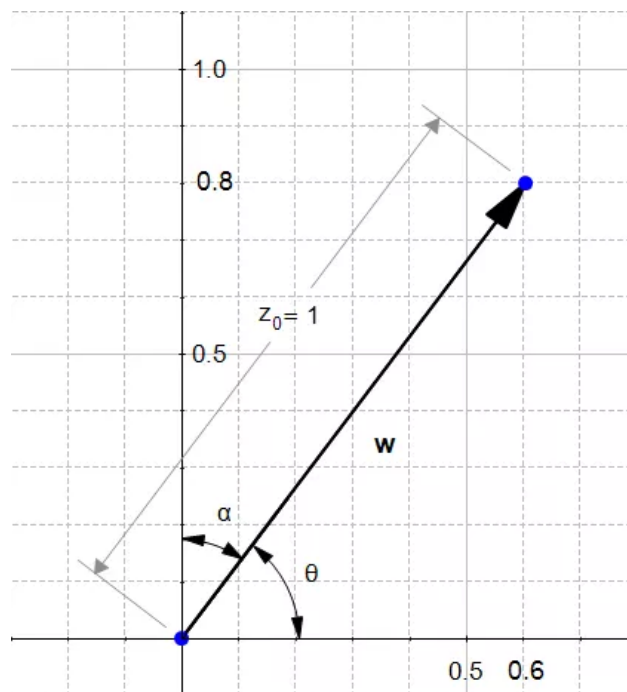*Figure 5: the direction of u*

We can see that **w** as indeed the same look as **u** except it is smaller. Something interesting about direction vectors like **w** is that their norm is equal to 1. That's why we often call them **unit vectors**.

The sum of two vectors

Figure 6: two vectors u and v
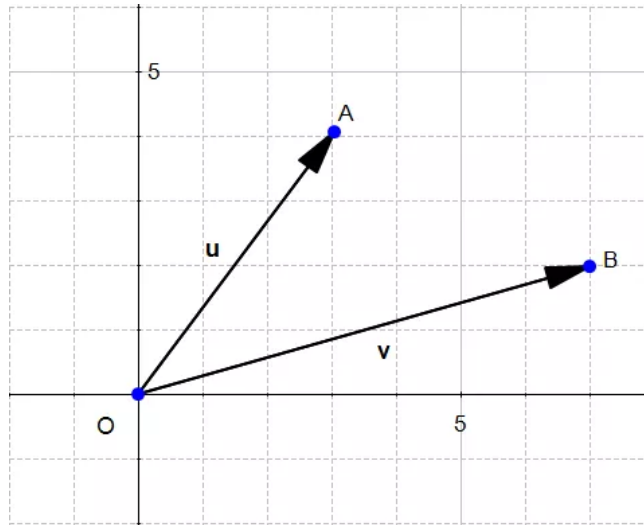
Given two vectors $\mathbf{u}(u_1, u_2)$ and $\mathbf{v}(v_1, v_2)$ then :

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2)$$

Which means that adding two vectors gives us **a third vector** whose coordinate are the sum of the coordinates of the original vectors.
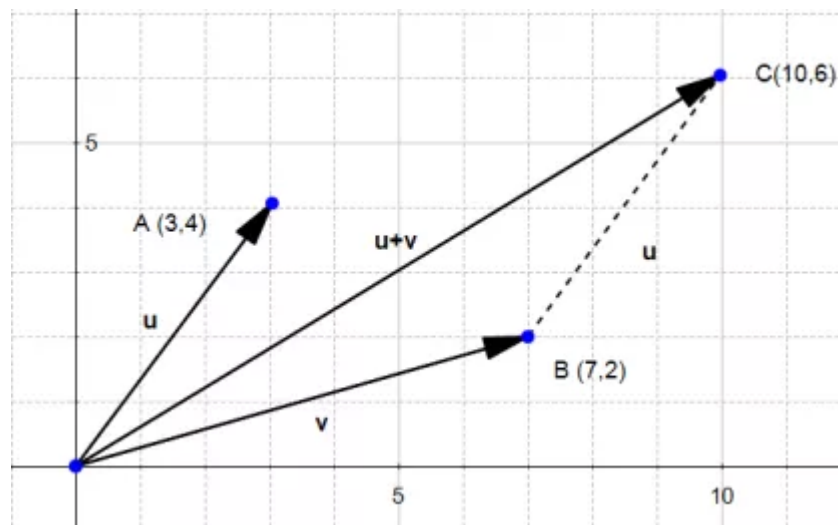
You can convince yourself with the example below:

Figure 7: the sum of two vectors

The difference between two vectors

The difference works the same way :

$$\mathbf{u} - \mathbf{v} = (u_1 - v_1, u_2 - v_2)$$
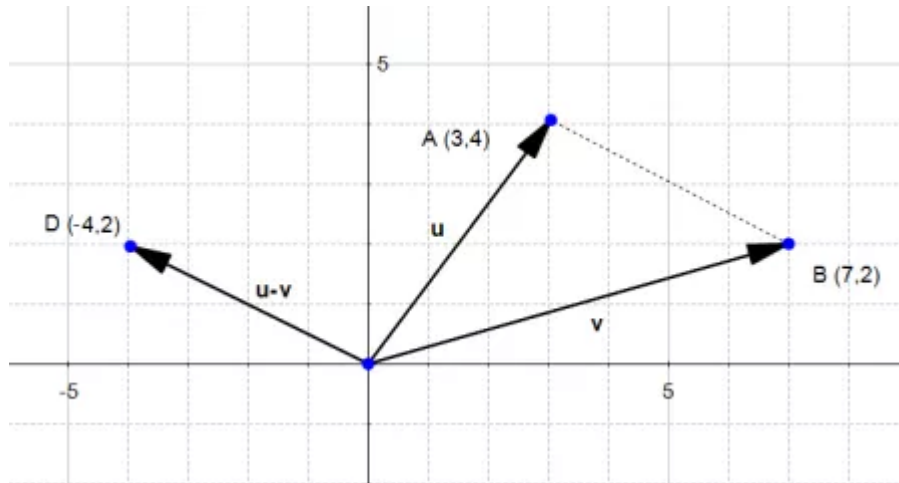
Figure 8: the difference of two vectors

Since the subtraction is not commutative, we can also consider the other case:

$$\mathbf{v} - \mathbf{u} = (v_1 - u_1, v_2 - u_2)$$

Figure 9: the difference v-u

The last two pictures describe the "*true*" vectors generated by the difference of $\mathbf{u}$ and $\mathbf{v}$.

However, since a vector has a magnitude and a direction, we often consider that parallel translate of a given vector (vectors with the same magnitude and direction but with a different origin) are the same vector, just drawn in a different place in space.

So don't be surprised if you meet the following :
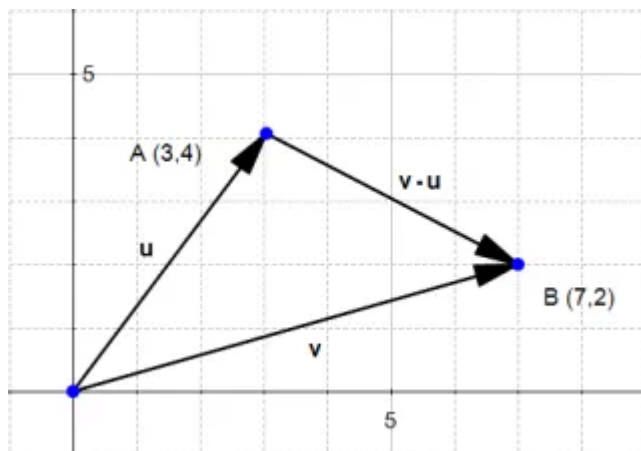
Figure 10: another way to view the difference v-u

and

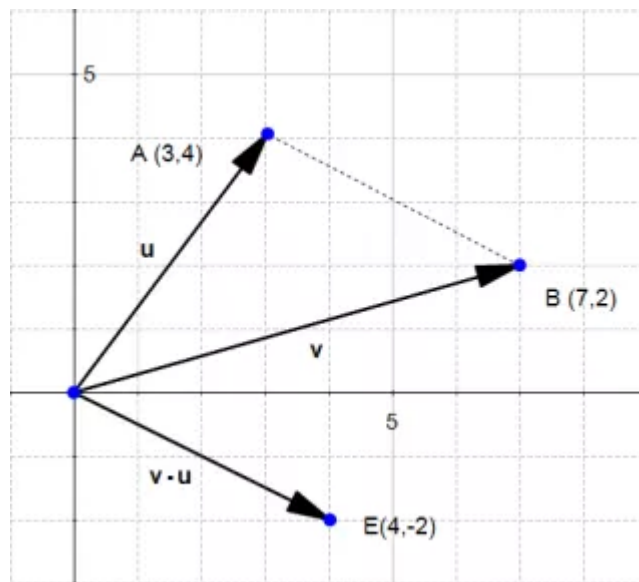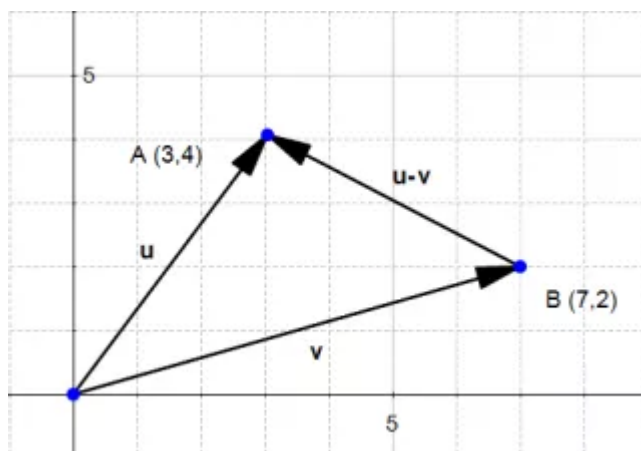*Figure 11: another way to view the difference u-v*

If you do the math, it looks wrong, because the end of the vector $\mathbf{u} - \mathbf{v}$ is not in the right point, but it is a convenient way of thinking about vectors which you'll encounter often.

The dot product

One **very** important notion to understand SVM is the dot product.

Definition: Geometrically, it is the product of the Euclidian magnitudes of the two vectors and the cosine of the angle between them

Which means if we have two vectors **x** and **y** and there is an angle θ (theta) between them, their dot product is :

$$\mathbf{x} \cdot \mathbf{y} = \|x\|\|y\|cos(\theta)$$

Why ?

To understand let's look at the problem geometrically.
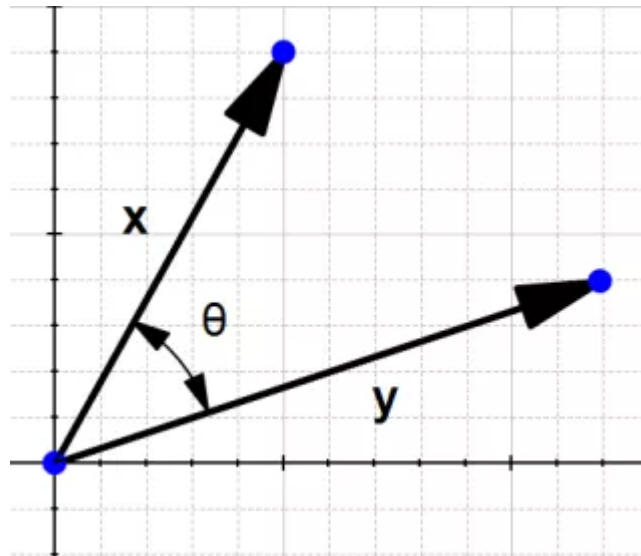
Figure 12

In the definition, they talk about $cos(\theta)$, let's see what it is.

By definition we know that in a right-angled triangle:

$$cos(\theta) = \frac{adjacent}{hypotenuse}$$

In our example, we don't have a right-angled triangle.

However if we take a different look Figure 12 we can find two right-angled triangles formed by each vector with the horizontal axis.
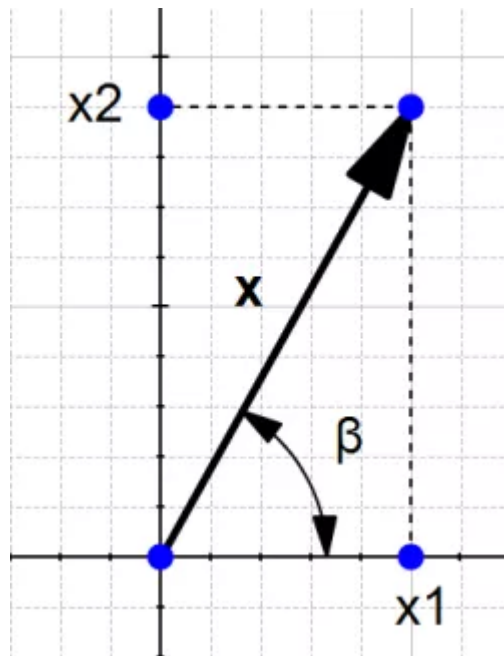
Figure 13

and
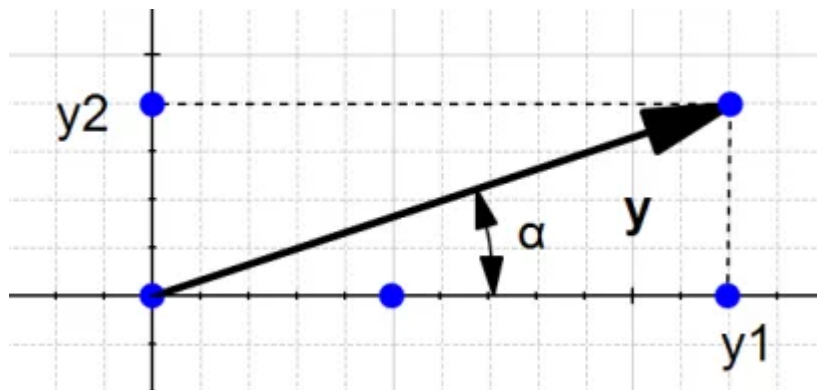
Figure 14

So now we can view our original schema like this:
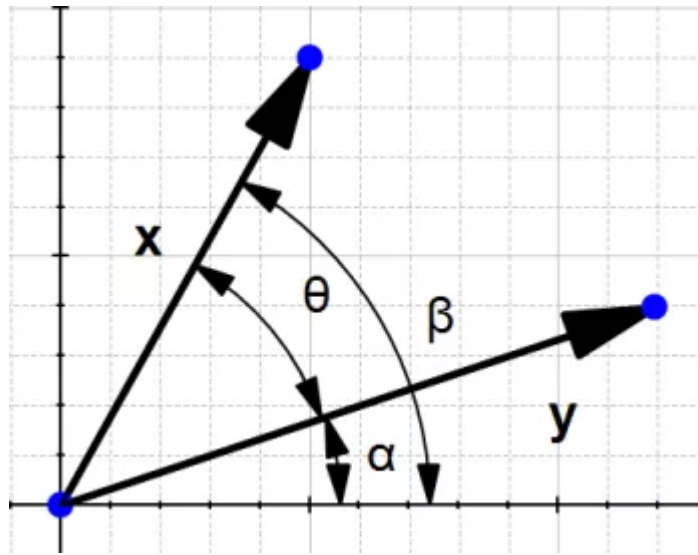
Figure 15

We can see that

$$\theta = \beta - \alpha$$

So computing $cos(\theta)$ is like computing $cos(\beta - \alpha)$

There is a special formula called the *difference identity for cosine* which says that:

$$cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha)$$

(if you want you can read the demonstration here)

Let's use this formula!

$$cos(\beta) = \frac{adjacent}{hypotenuse} = \frac{x_1}{\|x\|}$$

$$sin(\beta) = \frac{opposite}{hypotenuse} = \frac{x_2}{\|x\|}$$

$$cos(\alpha) = \frac{adjacent}{hypotenuse} = \frac{y_1}{\|y\|}$$

$$sin(\alpha) = \frac{opposite}{hypotenuse} = \frac{y_2}{\|y\|}$$

So if we replace each term

$$cos(\theta) = cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha)$$

$$cos(\theta) = \frac{x_1}{\|x\|}\frac{y_1}{\|y\|} + \frac{x_2}{\|x\|}\frac{y_2}{\|y\|}$$

$$cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{\|x\|\|y\|}$$

If we multiply both sides by $\|x\|\|y\|$ we get:

$$\|x\|\|y\|cos(\theta) = x_1 y_1 + x_2 y_2$$

Which is the same as :

$$\|x\|\|y\|cos(\theta) = \mathbf{x} \cdot \mathbf{y}$$

We just found the geometric definition of the dot product !

Eventually from the two last equations we can see that :

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 = \sum_{i=1}^{2}(x_i y_i)$$

This is the algebraic definition of the dot product !

A few words on notation

The dot product is called like that because we write a dot between the two vectors.
Talking about the dot product $\mathbf{x} \cdot \mathbf{y}$ is the same as talking about

- the **inner product** $\langle x, y \rangle$ (in linear algebra)
- **scalar product** because we take the product of two vectors and it returns a scalar (a real number)

The orthogonal projection of a vector

Given two vectors $\mathbf{x}$ and $\mathbf{y}$, we would like to find the orthogonal projection of $\mathbf{x}$ onto $\mathbf{y}$.
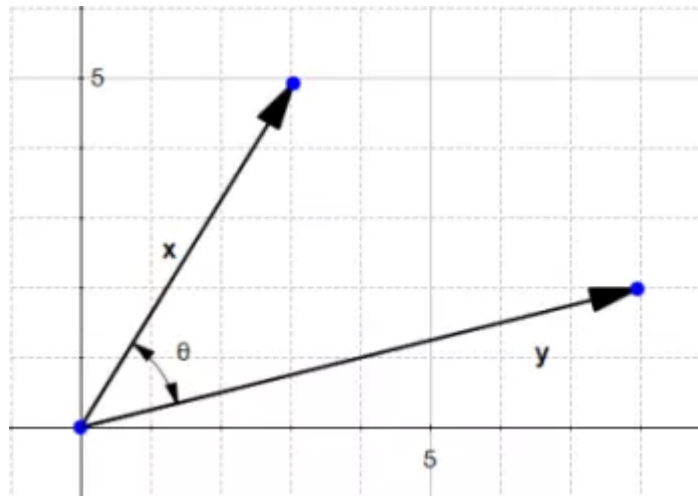
Figure 16

To do this we project the vector **x** onto **y**
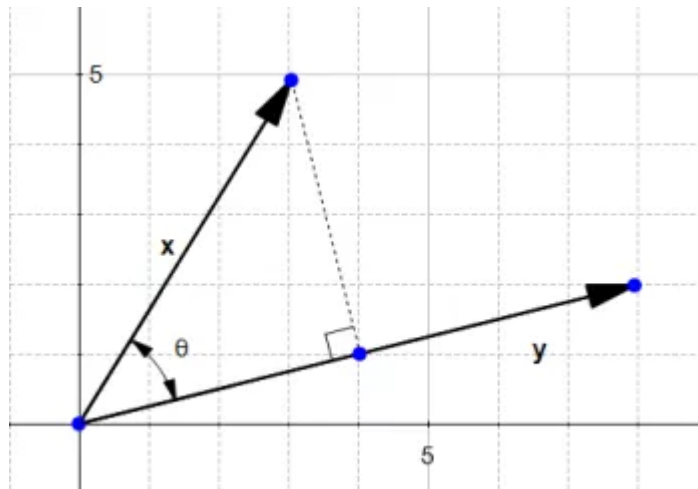
Figure 17

This give us the vector **z**
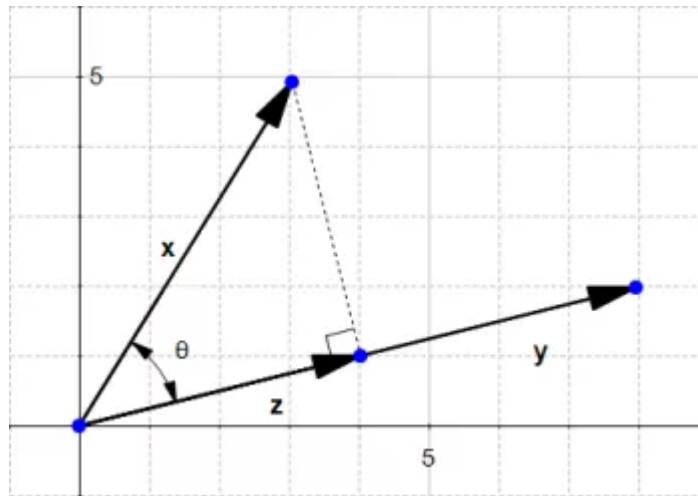
Figure 18 : z is the projection of x onto y

By definition :

$$cos(\theta) = \frac{\|z\|}{\|x\|}$$

$$\|z\| = \|x\|cos(\theta)$$

We saw in the section about the dot product that

$$cos(\theta) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|x\|\|y\|}$$

So we replace $cos(\theta)$ in our equation:

$$\|z\| = \|x\|\frac{\mathbf{x} \cdot \mathbf{y}}{\|x\|\|y\|}$$

$$\|z\| = \frac{\mathbf{x} \cdot \mathbf{y}}{\|y\|}$$

If we define the vector $\mathbf{u}$ as the **direction** of $\mathbf{y}$ then

$$\mathbf{u} = \frac{\mathbf{y}}{\|y\|}$$

and

$$\|z\| = \mathbf{u} \cdot \mathbf{x}$$

We now have a simple way to compute the norm of the vector **z**.
Since this vector is in the same direction as **y** it has the direction **u**

$$\mathbf{u} = \frac{\mathbf{z}}{\|z\|}$$

$$\mathbf{z} = \|z\|\mathbf{u}$$

And we can say :

The vector $\mathbf{z} = (\mathbf{u} \cdot \mathbf{x})\mathbf{u}$ *is the orthogonal projection of* $\mathbf{x}$ *onto* $\mathbf{y}$.

Why are we interested by the orthogonal projection ? Well in our example, it allows us to compute the distance between **x** and the line which goes through **y**.
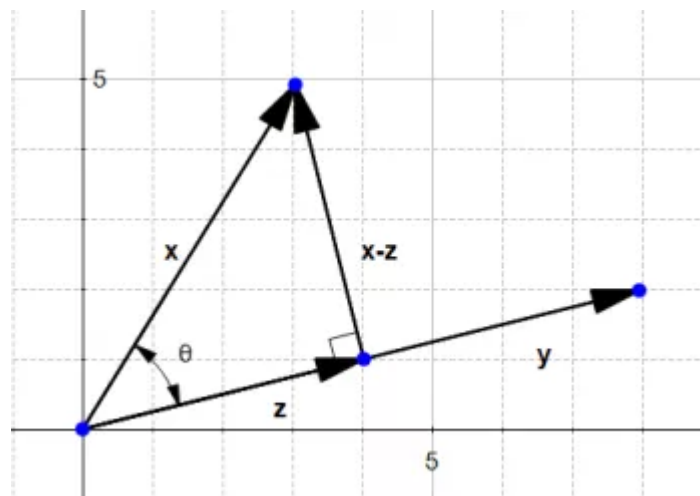
Figure 19

We see that this distance is $\|x - z\|$

$$\|x - z\| = \sqrt{(3-4)^2 + (5-1)^2} = \sqrt{17}$$

The SVM hyperplane

Understanding the equation of the hyperplane

You probably learnt that an equation of a line is : $y = ax + b$. However when reading about hyperplane, you will often find that the equation of an hyperplane is defined by :

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

How does these two forms relate ?

In the hyperplane equation you can see that the name of the variables are in bold. Which means that they are vectors ! Moreover, $\mathbf{w}^T\mathbf{x}$ is how we compute the inner product of two vectors, and if you recall, the inner product is just another name for the dot product !

Note that

$$y = ax + b$$

is the same thing as

$$y - ax - b = 0$$

Given two vectors $\mathbf{w} \begin{pmatrix} -b \\ -a \\ 1 \end{pmatrix}$ and $\mathbf{x} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$

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

$$\mathbf{w}^T\mathbf{x} = y - ax - b$$

The two equations are just different ways of expressing the same thing.

It is interesting to note that w_0 is $-b$, which means that this value determines the intersection of the line with the vertical axis.

Why do we use the hyperplane equation $\mathbf{w}^T\mathbf{x}$ instead of $y = ax + b$?

For two reasons:

- it is easier to work in more than two dimensions with this notation,
- the vector $\mathbf{w}$ will always be normal to the hyperplane(Note: I received a lot of questions about the last remark. $\mathbf{w}$ will always be normal because we use this vector to define the hyperplane, so by definition it will be normal. As you can see this page, when we define a hyperplane, we suppose that we have a vector that is orthogonal to the hyperplane)

And this last property will come in handy to compute the distance from a point to the hyperplane.

Compute the distance from a point to the hyperplane

In Figure 20 we have an hyperplane, which separates two group of data.
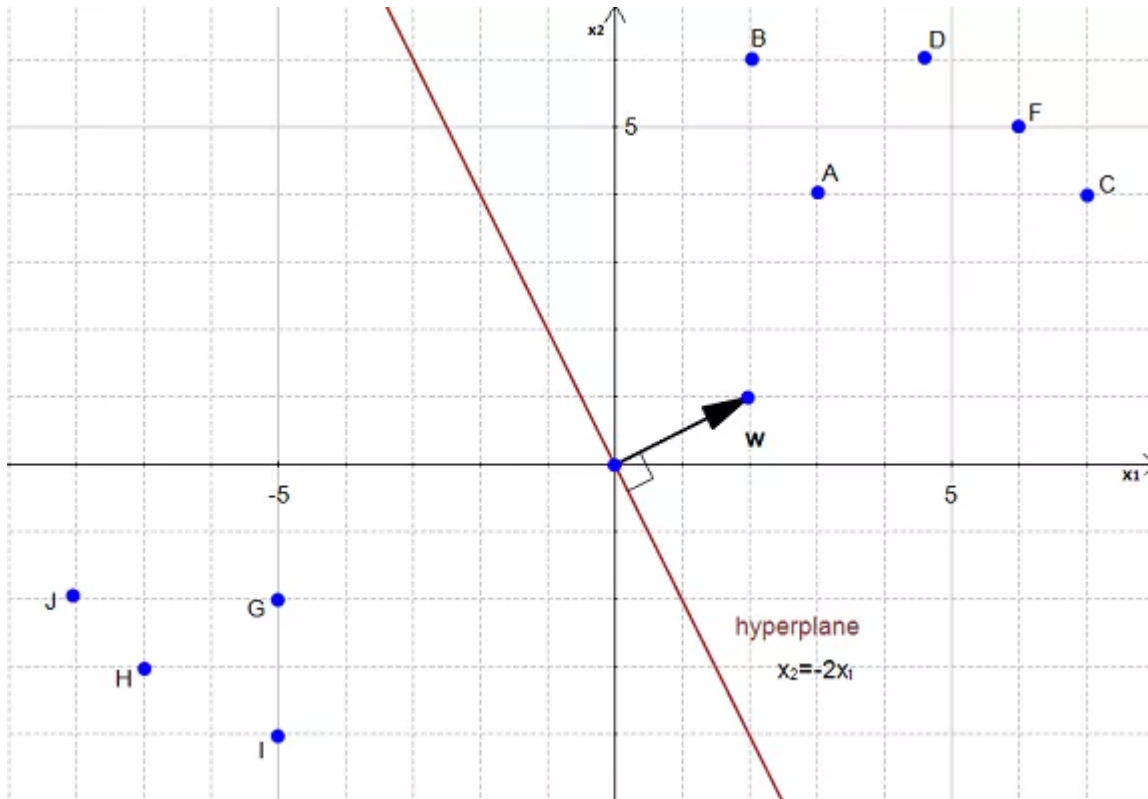
Figure 20

To simplify this example, we have set $w_0 = 0$.

As you can see on the Figure 20, the equation of the hyperplane is :

$$x_2 = -2x_1$$

which is equivalent to

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

with $\mathbf{w} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\mathbf{x} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

Note that the vector $\mathbf{w}$ is shown on the Figure 20. (w is not a data point)

We would like to compute the distance between the point $A(3,4)$ and the hyperplane.

This is the distance between A and its projection onto the hyperplane
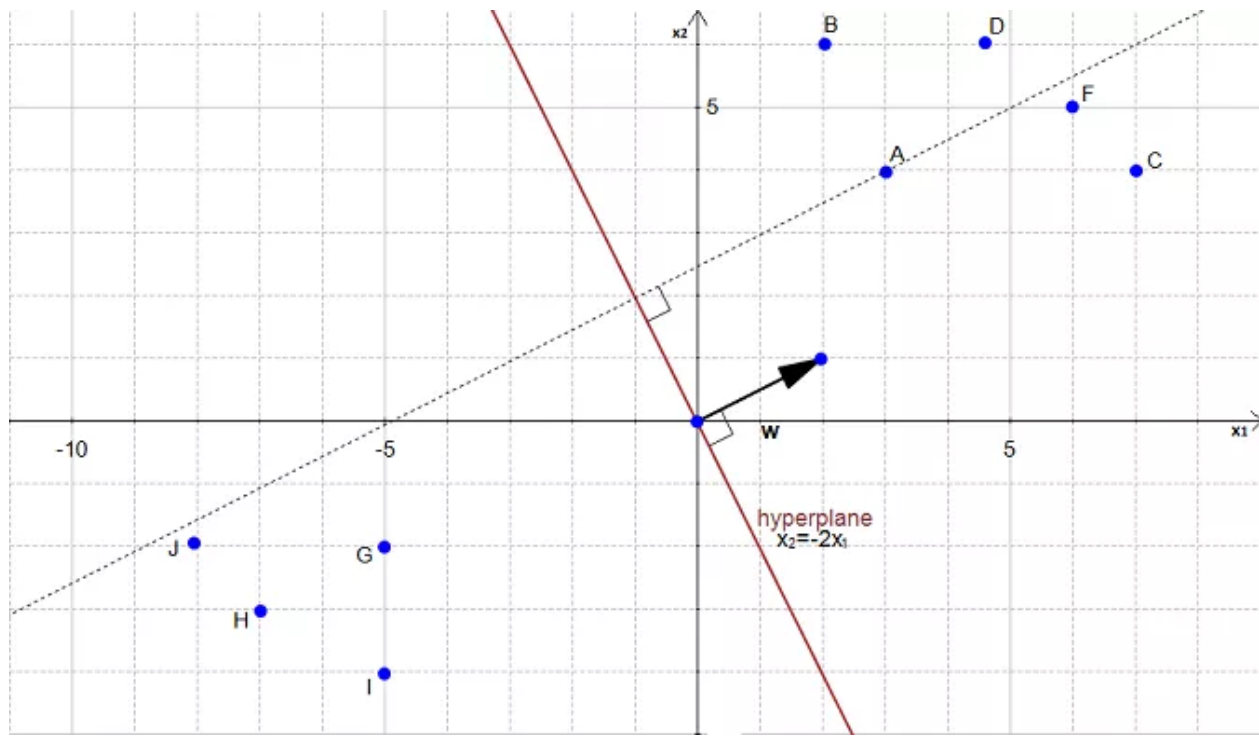
Figure 21

We can view the point A as a vector from the origin to A.
If we project it onto the normal vector $\mathbf{w}$
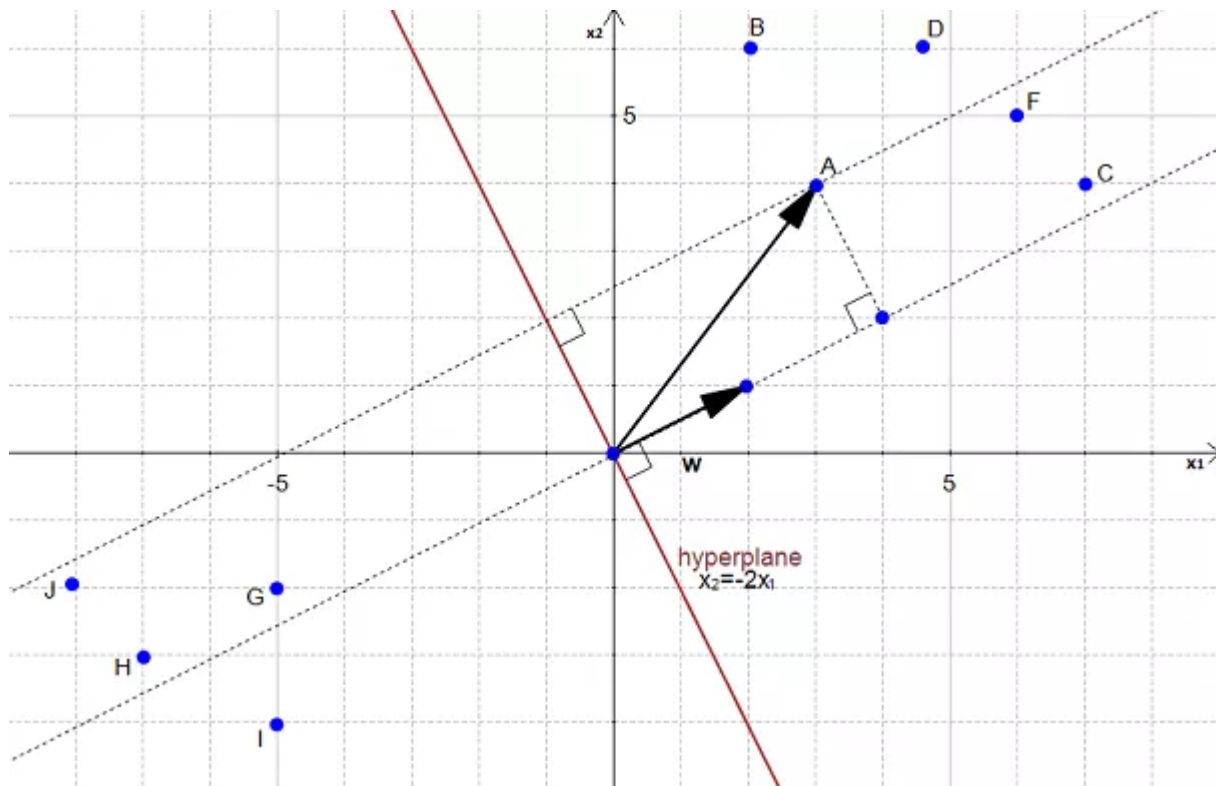
Figure 22 : projection of a onto w

We get the vector $\mathbf{p}$
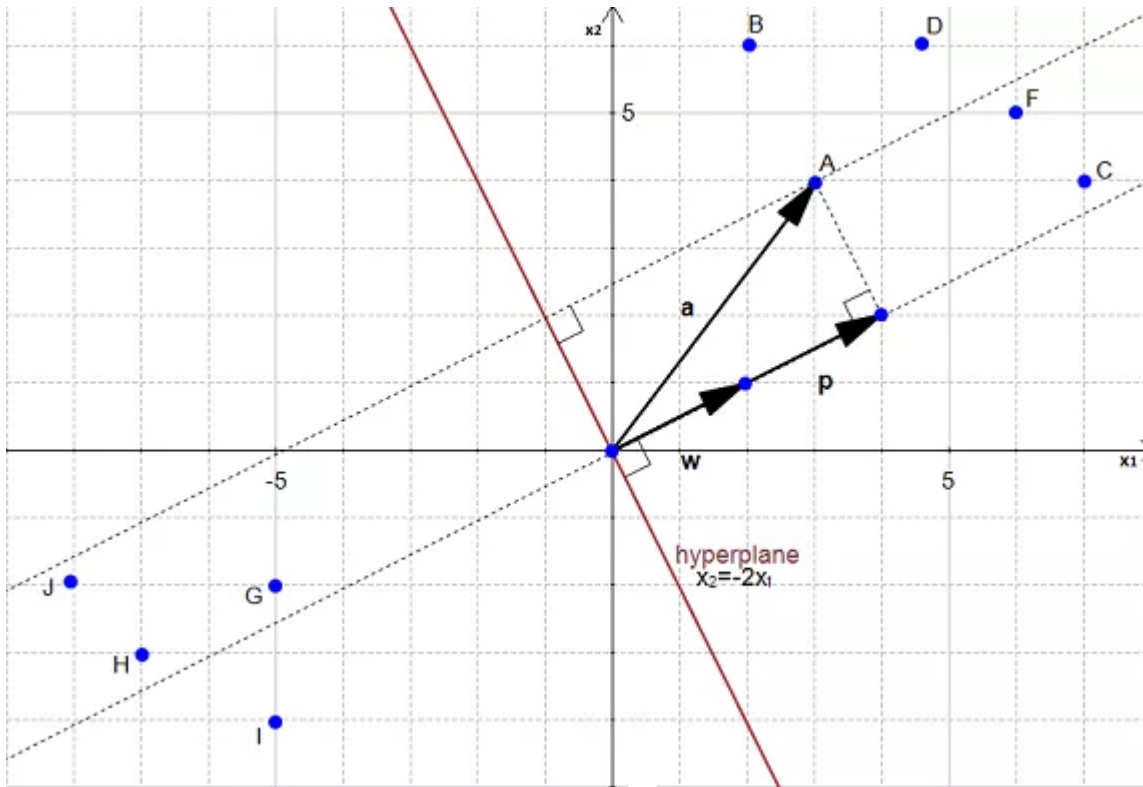
Figure 23: p is the projection of a onto w

Our goal is to find the distance between the point $A(3,4)$ and the hyperplane.
We can see in Figure 23 that this distance is the same thing as $\|p\|$.
Let's compute this value.

We start with two vectors, $\mathbf{w} = (2,1)$ which is normal to the hyperplane, and $\mathbf{a} = (3,4)$ which is the vector between the origin and A.

$$\|w\| = \sqrt{2^2 + 1^2} = \sqrt{5}$$

Let the vector $\mathbf{u}$ be the direction of $\mathbf{w}$

$$\mathbf{u} = (\frac{2}{\sqrt{5}}, \frac{1}{\sqrt{5}})$$

$\mathbf{p}$ is the orthogonal projection of $\mathbf{a}$ onto $\mathbf{w}$ so :

$$\mathbf{p} = (\mathbf{u} \cdot \mathbf{a})\mathbf{u}$$

$$\mathbf{p} = (3 \times \frac{2}{\sqrt{5}} + 4 \times \frac{1}{\sqrt{5}})\mathbf{u}$$

$$\mathbf{p} = (\frac{6}{\sqrt{5}} + \frac{4}{\sqrt{5}})\mathbf{u}$$

$$\mathbf{p} = \frac{10}{\sqrt{5}}\mathbf{u}$$

$$\mathbf{p} = (\frac{10}{\sqrt{5}} \times \frac{2}{\sqrt{5}}, \frac{10}{\sqrt{5}} \times \frac{1}{\sqrt{5}})$$

$$\mathbf{p} = (\frac{20}{5}, \frac{10}{5})$$

$$\mathbf{p} = (4, 2)$$

$$\|p\| = \sqrt{4^2 + 2^2} = 2\sqrt{5}$$

Compute the margin of the hyperplane

Now that we have the distance $\|p\|$ between A and the hyperplane, the margin is defined by :

$$margin = 2\|p\| = 4\sqrt{5}$$

We did it ! We computed the margin of the hyperplane !

Conclusion

This ends the Part 2 of this tutorial about the math behind SVM.
There was a lot more of math, but I hope you have been able to follow the article without problem.

What's next?

Now that we know how to compute the margin, we might want to know how to select the best hyperplane, this is described in Part 3 of the tutorial : How to find the optimal hyperplane ?

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 November 9, 2014 [https://www.svm-tutorial.com/2014/11/svm-understanding-math-part-2/] .

127 thoughts on "SVM - Understanding the math - Part 2"

Oleg Prutz
November 21, 2014 at 9:29 pm

Are you planning to tell about support vectors, non-linear kernels and optimization (I mean finding the minimum of the distance from the hyperplane to the suport vectors) in this tutorial? It seems that one need to know optimization theory in depth to understand this algorithm. It would be nice to see the simple explanation of what the algorithm is doing actually.

Alexandre KOWALCZYK Post author
November 22, 2014 at 12:32 pm

Yes that is what I am planning to do. However optimization theory is indeed very important to understand the algorithm and I am still figuring out how to explain SVM without going too deep into details.

Ranjith
January 19, 2017 at 4:55 pm

Hello Alexandre.. Great explanation.. Thanks..
I have a question on figure19.

My interpretation below:
We are assuming $\|x-z\|$ to be the hypotenuse but actually it is not the case right? the vector opposite to $\|x-z\|$ is hypotenuse (X). So the formula following that should substract the distance from the hypotenuse instead of adding.

I'm bit confused as I'm new to this. Kindly help me clarify my understanding.

Alexandre KOWALCZYK Post author

January 21, 2017 at 9:06 pm

No, we are not assuming ‖x−z‖ to be the hypotenuse.

When I do the calculation in figure, I say, let us compute the norm of the vector x-z. Let us call this vector k. What are the coordinates of k? k(5-1, 3-4) which gives us k(4,-1), then we compute the norm as usual ‖k‖ = sqrt(4^2+1^2) = sqrt(17)

We can find the same result by applying Pythagoras' theorem multiple times. Let us consider the hypotenuse x. The theorem tells us that ‖x‖^2 = ‖z‖^2 + ‖x−z‖^2. Which is equivalent to ‖x−z‖^2 = ‖x‖^2 - ‖z‖^2. ‖x‖^2 = 3^2 + 5^2 = 9 +25 = 34.
‖z‖^2 = 4^2 + 1^2 = 16 + 1 =17. If we replace= ‖x−z‖^2 = ‖x‖^2 - ‖z‖^2 = 34 - 17 = 17. Hence, ‖x−z‖ = sqrt(17).

ranjith

January 22, 2017 at 11:22 am

Very clear explanation again. Thank you.

Cecilia

March 30, 2017 at 7:27 am

Hi Alexandre, I have a question here. How to find the coordinates of Z?

Alexandre KOWALCZYK Post author

April 3, 2017 at 11:37 pm

The coordinate of z are (4,1). You can read them on the graph.

Bharath
April 7, 2018 at 8:56 pm

Hello alex. I wanna learn ml. Can you suggest me the process to learn it practically as you did. Can you suggest me books .

damon
July 23, 2017 at 12:38 am

hey, I am wondering in the real example, how do we know which data point is the closest to the hyperplane？By computing all the distances and compare them?

Franck Berthuit
December 11, 2014 at 5:43 pm

Very clear article, Alexandre... and enjoyable for a poor mathematician like me.
I'm eager to read then next one.
Bye

thewizardofme
March 16, 2015 at 12:37 pm

Marvellious article+lecture! Thanks for making it so clear.

thewizardofme
March 16, 2015 at 2:57 pm

Can you please explain non-linear SVMs and Kernel in your preceeding articles?

Alexandre KOWALCZYK Post author

March 19, 2015 at 12:09 am

Thanks for your kind comment. I need to find more time to write new articles. 🙂

Shivani Bhardwaj

April 28, 2015 at 3:35 am

I was trying to understand SVM from a very long time. your blog really helped me a lot and now I know what I am dealing with. your tutorial not only helped in understanding the mathematical jargon but also give me the clear perspective of what I am doing.
Thanks a lot!!

Alexandre KOWALCZYK Post author

April 29, 2015 at 11:52 pm

I am glad to hear it helped you. Thanks 🙂

Shyam

April 6, 2015 at 12:22 pm

Very lucid explanation - looking forward for part 3. When's that coming out ?

Alexandre KOWALCZYK Post author

April 8, 2015 at 8:20 pm

Thanks for the comment Shyam. I am afraid that recently I have spent most of my time on kaggle competitions and playing with convolutionnal neural networks. I will try to write the following part in the coming weeks in order to no achieve this work.

Nasser Tamim
November 26, 2016 at 12:01 am

thank you very much very your wonderful explanation
and i am waiting patiently for lectures in deep learning especially convolution neural network

Pragatheeswaran
August 2, 2017 at 11:22 am

Very detailed explanation! Great work! Please notify us when you publish tutorials on CNN.

korawit
April 8, 2015 at 8:54 am

This is the best tutorial ever!

Kunal
May 3, 2015 at 2:41 pm

Of all the links I found while doing a google search on SVM this is by far the best one in terms of simplicity of language in which it is explained...Thanks Alex

ajay
May 5, 2015 at 1:21 pm

Very Nice explanation. Where can I get part 3

Alexandre KOWALCZYK Post author
May 6, 2015 at 7:33 pm

I am currently writing it. But it is coming soon. 🙂

---

dragon518
May 19, 2015 at 11:14 pm

This is the best blog about SVM I have seen ever, help me so much, thank you very much, look forward to excellent part 3. BTW, "To simplify this example, we have set $w_0 = 0$", do you mean that setting the start point of vector $\mathbf{w}$ at origin?

Alexandre KOWALCZYK Post author
May 21, 2015 at 10:17 pm

Thanks for your kind comment. No this does not mean setting the start point of the vector $\mathbf{w}$ at the origin. We could place it somewhere else because we often consider that the parallel translate of a given vector is the same vector (this is illustrated in the section about the difference of two vectors)In the definition of the equation of a hyperplane the $\mathbf{w}$ vector is a 3-dimensional vector $\mathbf{w}(w_0, w_1, w_2)$. By setting w_0 to 0 we can do the remaining calculations with a 2-dimensional $\mathbf{w}(w_1, w_2)$ vector. Because the definition says that $w_0 = -b$ and we use $w_0 = 0$ instead, it removes the intercept term from the equation. As a result the hyperplane passes through the origin. In the Part 3 I wrote in more details about the hyperplane equation, things should be easier to understand.

Tam Dc
May 27, 2015 at 5:40 pm

Very helpful ! Thank you. Hope part 3

Farzad

June 2, 2015 at 9:07 am

Thank you!

Goutham

June 12, 2015 at 5:53 pm

Nice Explanation.Waiting for future posts 🙂

Alexandre KOWALCZYK Post author

June 14, 2015 at 5:30 pm

Thanks. The part 3 is now online. (I added the link at the end of the article)

David

June 20, 2015 at 5:34 pm

Thanks a lot, Alexandre!

Gabriel B. Théberge

July 6, 2015 at 9:34 pm

That is effectively crystal clear! I have read a lot of papers on this topic but nothing was as clear and accessible as your presentation Alexandre!

Felicia
September 30, 2015 at 7:16 pm

This is the most useful blog about SVM I've seen so far, especially for people like who don't have much knowledge in linear algebra.

A dumb question: why is the direction of \mathbf{w} perpendicular to the hyper-plane?

Alexandre KOWALCZYK [Post author]
October 2, 2015 at 10:01 pm

Thanks Felicia 🙂 For your question you can see my answer to the same question in the comments of part 3.

sidra
October 22, 2015 at 5:41 pm

it is a best lecture article .....
i have read many lectures but now the concept is clear .....

Ankit PK
October 23, 2015 at 11:25 am

Awesome Alexandre..... Thanks for this nice series...... makes SVM a lot easier

Subha MG
November 11, 2015 at 8:08 am

Hi Alexandre..Your blog is simply superb! The way you've explained concepts!! I saw several videos on SVMs..but I didn't get a clear picture..Your articles have made it super-clear!! Super-like!!

MIng
November 21, 2015 at 5:00 am

really helpful, fantastic

Md. Asadur Rahman
November 25, 2015 at 5:53 am

No Word to thank you, brother! I was very worried and eager to learn about SVM, You have solved my problem. Be blessed by Almighty.

Sameer Panna
December 9, 2015 at 4:52 pm

Great Blog.. and excuse my ignorance but can you please explain how one arrives here w(−b,−a,1) and x(1,x,y)

Alexandre KOWALCZYK `Post author`
December 13, 2015 at 9:48 pm

Thank you. You find these two vectors by continuing the reasoning.
We want to express the equation y-ax−b=0 with a dot product between two vectors.
The dot product is the sum of several products. In our case there is two minus signs, so there is three elements being summed together. Our vectors will have three elements each. Then we transform the equation to display these products: y−ax−b=0 is equivalent to y*1−a*x−b*1=0 and then we transform the differences into sums :
y*1+(−a)*x+(−b)*1=0

Sameer Panna
December 14, 2015 at 9:06 pm

Thank you 🙂

elmustafa1
January 5, 2016 at 3:56 am

This is extremely helpful. Thank you so much.

Koushik
January 8, 2016 at 12:47 pm

Wonderful post.....it gives me clear understanding even in some of Linear Algebra concepts. Thanks....keep this good work up..... 🙂

adhi21
January 18, 2016 at 1:40 am

I am understand now why the equation is w^(T) . x + b. I have another question, what does "T" mean in that equation?
Thank you

Alexandre KOWALCZYK Post author
January 18, 2016 at 10:59 pm

It is the transpose of a matrix. (It transforms a column vector into a row vector for instance)

omar
February 9, 2016 at 2:46 am

thank you very much about that useful tutorial , can you write an article about dealing SVM with non_linear dataset

Alexandre KOWALCZYK Post author
February 14, 2016 at 9:47 pm

Thanks for the suggestion. I will try to finish this tutorial series first. 😉

sawi
February 23, 2016 at 8:41 am

Thanks for a very useful article, explaining every tiny detail about the calculation with simple language and figures.

chadaphone
March 11, 2016 at 6:48 am

It is very helpful. Thank you very much. Great job!!

chansungpark
March 25, 2016 at 1:23 am

Thank you for very explicit explanation.
I have a question since I have no background knowledge about cosine, sine, etc.
Shouldn't Direction of vector be just angle of the triangle? I am just curious what cos(β)= adjacent/hypotenuse formula fundamentally means?

Alexandre KOWALCZYK Post author
April 3, 2016 at 8:21 am

No because you can have another vector with the same angle between the axis and itself but with the vectors pointing in another direction. By using cosine, we use the length of the adjacent and hypotenuse and as we are using coordinates we obtain a vector pointing in the same direction.

Ravi
April 3, 2016 at 4:01 am

How can we show that the vector w will always be normal to the hyperplane?

Alexandre KOWALCZYK Post author
April 3, 2016 at 8:17 am

This is by definition of the hyperplane. You can read more about it here.

Farah
April 28, 2016 at 9:38 pm

Can u plz tell, why u multiplied distance between A and hyperplane by 2 to compute margin?

Alexandre KOWALCZYK Post author
April 28, 2016 at 9:54 pm

This was a simplification. The margin is explained in more details in Part 3.

Brandon
May 15, 2016 at 11:29 am

Thank for you blog, that is great. However, i have a question about the W(-b,-a,1) and X(1,x,y),
the transpose of W is a column vector and X is a row vector, the result of [column * row] is a matrix that size is
(3,3), can you tell me where i missing?

Alexandre KOWALCZYK [Post author]
May 15, 2016 at 12:01 pm

Good catch. Indeed both w and x needs to be column vectors so that the transpose of w is a row and we do
[row * column]. I updated the article. Thanks!

Shawn
June 20, 2016 at 4:31 am

So far the best explanation of SVM in the net for those who do not have the required math background.
Fantastic job! Please post non-linear SVMs and Kernel explanation. Thanks a lot !

Ashutosh Srivastava
June 22, 2016 at 4:53 am

Very Nice and crystal clear explaination i have ever found on internet.
It will be very helpful if you give some practical demostration of how SVM and other
learning algorithms can be implemented and interpreted on various platforms like weka and orange.What is
confusion matrix and ROI. How that Wt-b equation is generated etc.
Giving practical demonstration will be very helpful.

Thanks and Regards...

Everaldo Aguiar
June 27, 2016 at 12:32 am

Phenomenal explanation of SVMs. Thanks a lot for taking the time to write and publish this. I was wondering if you would mind if I used brief excerpts of your content. I am preparing a few slides for a course that I will be teaching and found some of your images and explanations very helpful. I'll be sure to include citations and a reference to your blog posts.

Alexandre KOWALCZYK [Post author]
June 28, 2016 at 9:15 pm

Thanks for your comment. No problem you can use some excerpts. For which course is it?

Farai Leboho
August 1, 2016 at 1:33 am

Speechless, this is downright simple to understand. This makes SVM move from very hard to simply understandably, thanks a lot mate. At least now i have an idea of what's happening behind the scenes of svm.SVC().fit(),
Great work.

rssoni
August 11, 2016 at 2:19 pm

The best explanation I can think of. I made the concept clear. The writing style is lucid and understandable. Nothing else can be easier than this explanation of SVM. Awesome, thanks Alex

A Logical Geek

September 8, 2016 at 4:12 pm

You are amazing. Thanks a lot for this.. Because of lack of enough maths background i have having difficulty reaching here . You helped a lot. Is it possible for you to explain the justification of langrange's multipliers as well as further explanation of SVM.

leviliard
September 14, 2016 at 9:50 pm

Thank you very much. Your a are a big teacher.

Ganecian
September 16, 2016 at 5:48 am

I'm still confused in determining normal vector w. If the equation of the hyperplane is $x2 = 1/3 * x1 + 1$, what is the normal vector w? How to calculate it when w0 is not zero. Thanks

Alexandre KOWALCZYK Post author
September 17, 2016 at 7:38 pm

In your example: $x2 = 1/3 * x1 + 1$, is in the form: $x2 = a * x1 + b$. To get the normal vector you just get the vector w(a,-1). So in this case, we define w(1/3,-1), x(x1,x2) and b = 1. And you can see that wx+b=0 is equivalent to $x2 = 1/3 * x1 + 1$.

Now if we plot, the vector w(1/3,-1), we can start to draw it where we want. I could start drawing at the origin x(0,0) or I can start drawing it directly on the hyperplane. I choose to do so, and I start drawing it at x(1,1+1/3). As you can see in the figure: it is normal to the hyperplane.

W is normal

Where I chose to start drawing it, does not change the fact that it is normal to the hyperplane.

Beungeut Boloho
September 17, 2016 at 11:17 am

Hi, can you explain what is the bias b visually? Is it the distance of vertical axis to the origin or the distance of the hyperplane to the origin?

Alexandre KOWALCZYK [Post author]
September 17, 2016 at 8:10 pm

Given an hyperplane having the equation wx+b=0 with vectors w(w0,w1) and x(x0,x1). b is the distance between the vertical axis and the origin only when the value w1 of the weight vector is equal -1. Indeed, when we transform this hyperplane equation to a line equation of the form y=ax+c we get a = -w0/w1 and c = -b/w1. Some books represent b as being the distance between the origin and the hyperplane, but I think this is true only under certain conditions, at least that is what I found when trying to verify it by myself using the first formula of this article using formulas from this page.

Srinath Shiv Kumar
September 26, 2016 at 5:25 am

Very clearly explained. Thanks a lot!

Bahareh Moradi
September 28, 2016 at 7:48 am

Thank you a thousand times…………You explained Lagrange multipliers in the best way in the world…..
can you introduce me some useful books which I can read and get more information about classification?

Sara Q. Abedulridha
October 22, 2016 at 10:31 pm

Many thanks brother, that is great…
wishing you all good things, God bless you…

Jeetendra Ahuja
October 28, 2016 at 6:42 am

Awesome tutorial, TAL man!
Just a small suggestion, when you give a link like for "cumulative", "dot product" , can you change a code of your site such that after clicking on this link, it gets open in different tab instead of opening in current tab.

Alexandre KOWALCZYK   Post author
October 31, 2016 at 6:05 pm

Hello. Thank you for your comment. I thought all my links were opening in a new tab but indeed it was not the case. I updated all the problematic links in this article. Thank for the remark !

Sonia
November 7, 2016 at 7:43 am

Sir,

x+.w+b= +1
x_.w+b = -1
why it is always equal to +1 and -1 for positive and negative support vector respectively.?. How to normalize this distance of hyperplane to support vectors?. Once we normalize, it always remains same for any kind of data. could you explain?. I am not clear about the distance between hyperplane and support vectors

---

Alexandre KOWALCZYK   Post author
November 15, 2016 at 10:22 pm

It is always equal to +1 and -1 because we are free to select w for which it will be the case (we can rescale w and b and keep the same hyperplane). So we decide arbitrarily to select among the ones for which it is equal to +1 and -1 because it will make the following computation easier.

---

tkotresh
November 16, 2016 at 6:18 pm

Thank you for finding time to write this article. I am back to basics and enjoying it as well.

AnasA. Hadi
November 18, 2016 at 1:36 am

Thanks for this illustration about SVM..

Vijay Gupta
January 12, 2017 at 7:04 pm

Hi ... Very good and detailed article . But one doubt : the vector w will always be normal to the hyperplane ..
Why is this so ?

Alexandre KOWALCZYK [Post author]
January 15, 2017 at 5:51 pm

Hi. This is by definition of the hyperplane. 🙂

Hamala
February 11, 2017 at 3:44 am

Hi, can I get a pdf version of this tutorial or a printable version 🙂 cheers thank you!

Alexandre KOWALCZYK [Post author]
February 15, 2017 at 10:58 pm

Hi, I thought that it was pretty easily printable with the theme I used. Do you mean you would like a printable version of the whole tutorial? Do you have trouble printing an invidual article?

YAN KANG
February 8, 2017 at 12:12 am

Yes, I have doubt here too. What does the normal mean here? I think it mean Orthogonal since two vectors u and v whose dot product is u·v=0 are said to be orthogonal.

---

siddhant
February 14, 2017 at 7:12 am

I had taken machine learning as my final semester research topic.For about a month I was unable to decide which topic to specifically decide to work upon.After a month my guide told me to work upon SVM in image processing.I had a little knowledge about SVM but the math part was very difficult.It was intricate as to say. Finally I came upon this blog and found it very help , the math was really very detailed yet simple to understand.Thanks for such a nice article the 6 part series has helped me a lot in understanding machine learning as a whole.

---

ridarafisyed
February 19, 2017 at 7:30 am

very very nice and clear article. but please will you tell me about "the vector w will always be normal to the hyperplane". i don't understand this point...

Alexandre KOWALCZYK [Post author]
February 23, 2017 at 9:32 pm

As stated in the comments before, this is by definition. We defined the hyperplane with this equation, so w is normal. If you read the example in this page: you can see that they say "Let's also suppose that we have a vector that is orthogonal (perpendicular) to the plane".

ridarafisyed
February 19, 2017 at 7:55 am

awesome article finally i understand fancy W T X = 0 and other stuff thanks sir for such a nice lecture!

pradeep
March 22, 2017 at 7:02 am

We can see that ww as indeed the same look as uu except it is smaller. Something interesting about direction vectors like ww is that their norm is equal to 1. That's why we often call them unit vectors. -- why direction vectors like 'w' is that their norm is equal to 1? can you please explain?

Nick
April 19, 2017 at 4:28 pm

I understood the proceeding for calculate the distance but there is a method more simple for do it: distance= dot(A,w)/||w|| Is there a particular reason for prefer your described method rather than my simpler?

Great tutorial! Compliments!

Alexandre KOWALCZYK [Post author]

April 19, 2017 at 10:46 pm

Thank you for your comment 🙂

Yes indeed your method is simpler. And it is basically what we are doing in the article because dot(A,w)/||w|| = dot(A,u) and then we create p using dot(A,u)u.

My goal was to use the projection p so that you can **see** what is this distance in the figure.

Al Fritsch
April 25, 2017 at 1:31 am

Question for you, why did you call the resulting vector "u" instead of calling it "y" hat which it really is (a unit vector in the direction y).

Alexandre KOWALCZYK Post author
April 27, 2017 at 12:55 pm

I believe we can call the vector the name we want as long as we define it explicitly.

Jill Minor
April 28, 2017 at 5:15 pm

I can't thank you enough for this. I'm in 500 level Artificial intelligence but I have no real background, even in linear algebra. You saved me.

Rahul
May 3, 2017 at 6:19 am

Hi, great tutorial I am learning lots from this. I have a doubt though, to find the orth proj of a on X1=-2x2 you have used w(1,2), how did u arrive to this value of w? The ortho project of a onto hyper plane X1=-2x2 should be same from any point normal to it, but your equations are dependent on the values of w and if I were to choose a different line normal to hyperplane I would get a diff magnitude won't I ? It would be great if you can explain

Alexandre KOWALCZYK Post author

May 14, 2017 at 6:23 pm

Yes indeed but it does not matter if you get a different magnitude because you are trying to minimize the distance and you would use the same w in all your calculations. This ability to choose the w arbitrarily is used later when we define the optimization problem. I explain this in detail in my ebook.

Sunder

May 10, 2017 at 1:50 pm

Hi,

It is showing this error in many places on the page: '[Math Processing Error]'

Can you please check?

Regards,
Sunder

Sunder

May 10, 2017 at 2:27 pm

Now, it looks fine. Don't knw what caused it last time.

Thanks for the prompt reply.

Sunder

uday
May 25, 2017 at 9:04 am

"" Our goal is to find the distance between the point A(3,4) and the hyperplane.
We can see in Figure 23 that this distance is the same thing as ‖p‖""

How did you know that the distance of to hyperplane is equal to ‖p‖ ??

Alexandre KOWALCZYK [Post author]
May 27, 2017 at 7:17 pm

Start from the hyperplane and follow the dotted line until you reach A. This is the distance we want to find.
Start from the hyperplane (0,0) and follow the line until you reach the end of p. You can see that this distance is the same. You have to see that p is the vector starting at (0,0 and finishing at (4,2). ||p|| the length of p, so ||p|| is this distance we are looking for

Lalit
May 29, 2017 at 9:56 am

This really easy to understand. I had forgotten concepts of vectors, It came handy as a refresher.

rio
June 13, 2017 at 4:24 pm

I saw a lot of materials related to SVM…
(Almost materials are not good to understand.. because
explanation is not good I think..)

BUT THIS MATERIAL IS THE BEST EXPLANATION!!!

very very thank you !!!!

Pratik
July 10, 2017 at 3:03 pm

Hi Alex,

All the mathematics and calculus would be more clear if you can show it with some working example.
Example: We have a data set that has 2 explanatory variables i.e x1 and x2 and one binary outcome variable y.
Then how would we calculate hyper plane and w values.

Pragatheeswaran
August 2, 2017 at 11:43 am

A small doubt,
Before fig 19, The formula to find the vector is derived as,
z = (u.x)u
and in the derivation I found this equation
|z| = u.x ------------(remember)

After fig 23 to find p, you said p = (u.a)u
so like that we can find |p| = u.a using this

but you are trying to find vector p and then |p|
so if our ultimate aim is to find |p| then we can use only |p|=u.a
without finding the actual vector

Alexandre KOWALCZYK [Post author]
August 7, 2017 at 10:09 pm

Yes indeed, we do not need to find the actual vector p to compute its norm, but I found it more easy to understand this way. Thanks for your insight!

janne
September 4, 2017 at 10:05 pm

Thank you for this great tutorial!
I have recently started my hopeless journey to understanding SVM. I'm wondering why this example starts hyperplane calculations from point origin (0,0)? how do equations change in situation when every data point is of the right side of (0,0). Where do we get the starting point?

Alexandre KOWALCZYK Post author
September 5, 2017 at 9:21 pm

Hello Janne and thank you for your comment. This example use an hyperplane passing through the origin in order to be more simple and avoid adding another term in the equation. 🙂

Raj
September 7, 2017 at 8:41 am

Your explanation is very helpful to me, but I have a doubt,
If we define the vector u

as the direction of y

then

u= y / ‖y‖ how we got this, please explain

Alexandre KOWALCZYK Post author
September 14, 2017 at 10:35 pm

We are using the definition of the direction: the direction of a vector x(x_1, x_2) is the vector w(x_1/|x|,x_2/|x|)

tuyen
October 4, 2017 at 4:36 am

Outstanding work! This is truly above and beyond. Keep looking for next publication

Lakshmikanta Nath
October 22, 2017 at 4:28 pm

hii...first of all i would like to thank you for sharing a wonderful site like this. I found this article quite interesting and useful for me.But, unfortunately I would like you bring your kind attention that, in this article, it is showing "MATH PROCESSING ERROR" instead of showing the actual math equation or any numeric.

Hope you will consider this situation and would make the necessary corrections and would reply me.

Have a nice day.

Thanking you.

Alexandre KOWALCZYK Post author
October 22, 2017 at 8:59 pm

Hello, Thank you for your message. The equations are displayed correctly on my computer. Maybe you have a problem with your web browser. I am using Chrome and everything looks good. You could try with another browser.

Lakshmikanta
October 23, 2017 at 1:15 pm

YES!! It worked...thanks again.

Shadab Faiz
November 5, 2017 at 2:17 pm

Nice article..!!! However 1 thing bug me most. Since we wanted to find the distance of A from hyperplane, then why don't we just use pythagoras theorum instead. That seems to be much easier.

Alexandre KOWALCZYK Post author
November 5, 2017 at 5:58 pm

We need to know the norm of p to use Pythagoras, so it seems the same for me. What was your idea?

GHALI
November 11, 2017 at 11:54 am

I want to thank you very very very much.
it's one of my best tutorial that i read and understand in my career.
thank's again.
my best reguard
have a nice day

Alexandre KOWALCZYK Post author
November 11, 2017 at 10:13 pm

Thanks a lot!

Erfan
November 29, 2017 at 6:56 pm

You are amazing Alexander , thank you very much. you did a big job !

god bless you.

i wish if you post a lecture about LDA and PCA .

thanks again

ateeq
December 1, 2017 at 7:49 pm

thanks you very much it is very helpful for me because t is explain well.

Anuj Gupta
December 10, 2017 at 5:02 pm

I am sorry but I am still confused about vector. I understand the math and the numbers you have shown here, but I cannot visualise vector yet in real world. When you add 2 and 3 apples I can visualise 5 apples. When you add two vectors, subtract them, dot, cross product, I understand the math, but I still cant visualise it.

Alexandre KOWALCZYK [Post author]
December 11, 2017 at 10:15 pm

Well, in figure 7 you can visualize that the sum of the two vectors (the two arrows) produce a third vector (the middle arrow). Here the arrow, is a representation of the vector. The same way I could represent the number one with the picture of one apple, and the number five with the picture of five apples. What is important to understand is that both a number and a vector should be considered as mathematical objects. And that these object have some properties. I can recommend you to read the books Mathematics a very short introduction and How to study as a mathematics major to get a better understanding of how you can consider mathematical objects.

ASHOKE SAHA
January 7, 2018 at 5:51 am

A very easy and fluent explanation of dot product, equation of plane and matrix combined. And the figures make it so clear. Thank you.!

regards

ashoke saha

Rathod Praveen Kumar
January 30, 2018 at 7:38 am

Thank you so much for such a beautiful explanation.

Adrian
February 24, 2018 at 12:49 pm

We need more people like you in this world! Clearer than any other resource I have seen out there. Please keep doing more great work like these 🙂

Sasha
April 7, 2018 at 2:20 pm

Hello, can you some explain. After the word "Define the vector u as the direction of y"
I dont understand how that formula appear u=y/|y| can anyone explain please why devide
I try find information but I can't in this case. Thanks

---

nourdine
May 16, 2018 at 12:09 pm

hi,
u is a unit vector ‖u‖=1, its direction is the same as y

valekar

May 16, 2018 at 11:13 am

Great Explanation!! Crisp and clear.

I just had one doubt, how did you consider the point "w" value to be (1,2) as the value? Is it chosen randomly?

Alexandre KOWALCZYK Post author

May 16, 2018 at 9:34 pm

Hello.

The value of w is derived from the equation of the hyperplane. Search the sentence starting with "As you can see on the Figure 20".