NMK40403 ARTIFICIAL INTELLIGENCE

SUPPORT VECTOR MACHINES (A)

Mohamed Elshaikh

Introduction

- One of the most performant off-the-shelf supervised machine learning algorithms.
- This means that when you have a problem and you try to run a SVM on it, you will often get pretty good results without many tweaks. Despite this, because it is based on a strong mathematical background, it is often seen as a black box.
- The first SVM algorithm is attributed to Vladimir Vapnik in 1963. He later worked closely with Alexey Chervonenkis on what is known as the VC theory, which attempts to explain the learning process from a statistical point of view, and they both contributed greatly to the SVM. You can find a very detailed history of SVMs here.
- In real life, SVMs have been successfully used in three main areas: text categorization, image recognition, and bioinformatics (Cristianini & Shawe-Taylor, 2000). Specific examples include classifying news stories, handwritten digit recognition, and cancer tissue samples.

What exactly is SVM ?

SVM is a supervised learning model

- It means you need a dataset which has been ***labeled***.

- ***Example***: I have a business and I receive a lot of emails from customers every day. Some of these emails are complaints and should be answered very quickly. I would like a way to identify them quickly so that I answer these email in priority.

- ***Approach 1***: I can create a label in gmail using keywords, for instance "urgent", "complaint", "help"

- The drawback of this method is that I need to think of all potential keywords that some angry users might use, and I will probably miss some of them. Over time, my keyword list will probably become very messy and it will be hard to maintain.

What exactly is SVM ?

- ***Approach 2***: I can use a supervised machine learning algorithm.

- Step 1: I need a lot of emails, the more the better.
- Step 2: I will read the title of each email and classify it by saying "it is a complaint" or "it is not a complaint". It put a ***label*** on each email.
- Step 3: I will ***train*** a model on this dataset
- Step 4: I will assess the quality of the prediction (using cross validation)
- Step 5: I will use this model to ***predict*** if an email is a complaint or not.

- In this case, if I have trained the model with a lot of emails then it will perform well. SVM is just one among many models you can use to learn from this data and make predictions.

- Note that the ***crucial*** part is Step 2. If you give SVM ***unlabeled*** emails, then it can do nothing.

What exactly is SVM ?

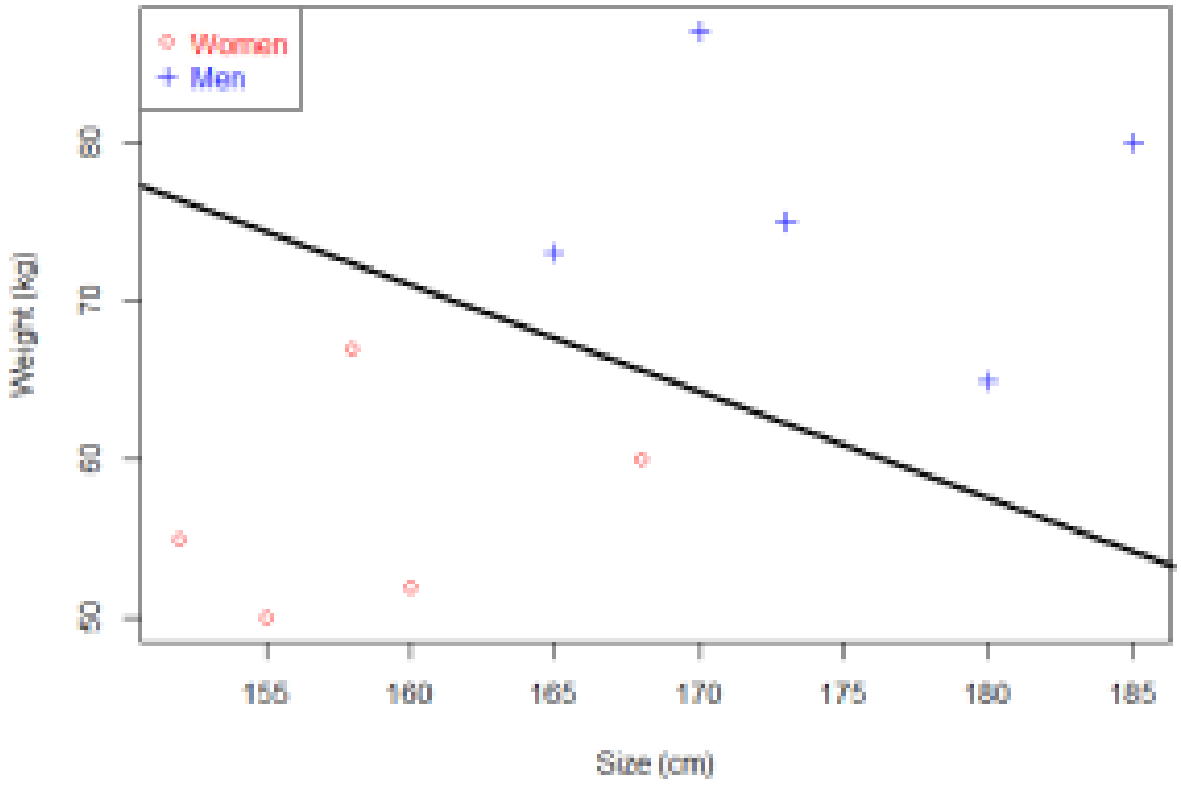
SVM learns a linear model

- Now we saw in our previous example that at the Step 3 a supervised learning algorithm such as SVM is trained with the labeled data. But what is it trained for? It is ***trained*** to learn something.
- What does it learn?

- In the case of SVM, it learns a ***linear model***.

- What is a linear model? In simple words: it is a line (in complicated words it is a hyperplane).
- If your data is very simple and only has two dimensions, then the SVM will learn a line which will be able to separate the data.

What exactly is SVM ?



The SVM is able to find a line which separates the data

- If it is just a line, why do we talk about a linear ***model***?
- Because you cannot learn a ***line***.

What exactly is SVM ?

- So instead of that:

  1) We suppose that the data we want to classify can be separated by a line

  2) We know that a line can be represented by the equation ***y=wx+b*** (this is our model)

  3) We know that there is an infinity of possible lines obtained by changing the value of ***w*** and ***b***

  4) We use an algorithm to determine which are the values of ***w*** and ***b*** giving the "best" line separating the data.

- SVM is one of these algorithms.

SVM - Understanding the math - Part 1 - The margin

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

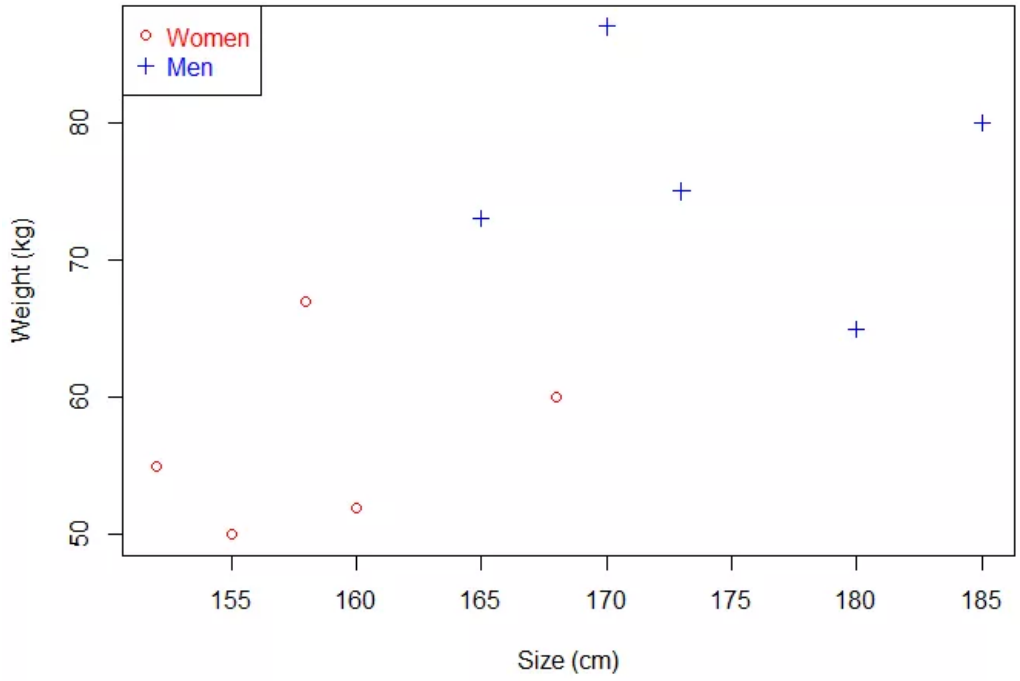
The goal of a support vector machine is to find the optimal separating hyperplane which maximizes the margin of the training data.

- The first thing we can see from this definition, is that a SVM needs training data. Which means it is a supervised learning algorithm.
- It is also important to know that SVM is a classification algorithm. Which means we will use it to predict if something belongs to a particular class.

SVM - Understanding the math - Part 1 - The margin

- For instance, we can have the training data below:

- We have plotted the size and weight of several people, and there is also a way to distinguish between men and women.
- With such data, using a SVM will allow us to answer the following question:
  - *Given a particular data point (weight and size), is the person a man or a woman ?*
- For instance: if someone measures 175 cm and weights 80 kg, is it a man of a woman?

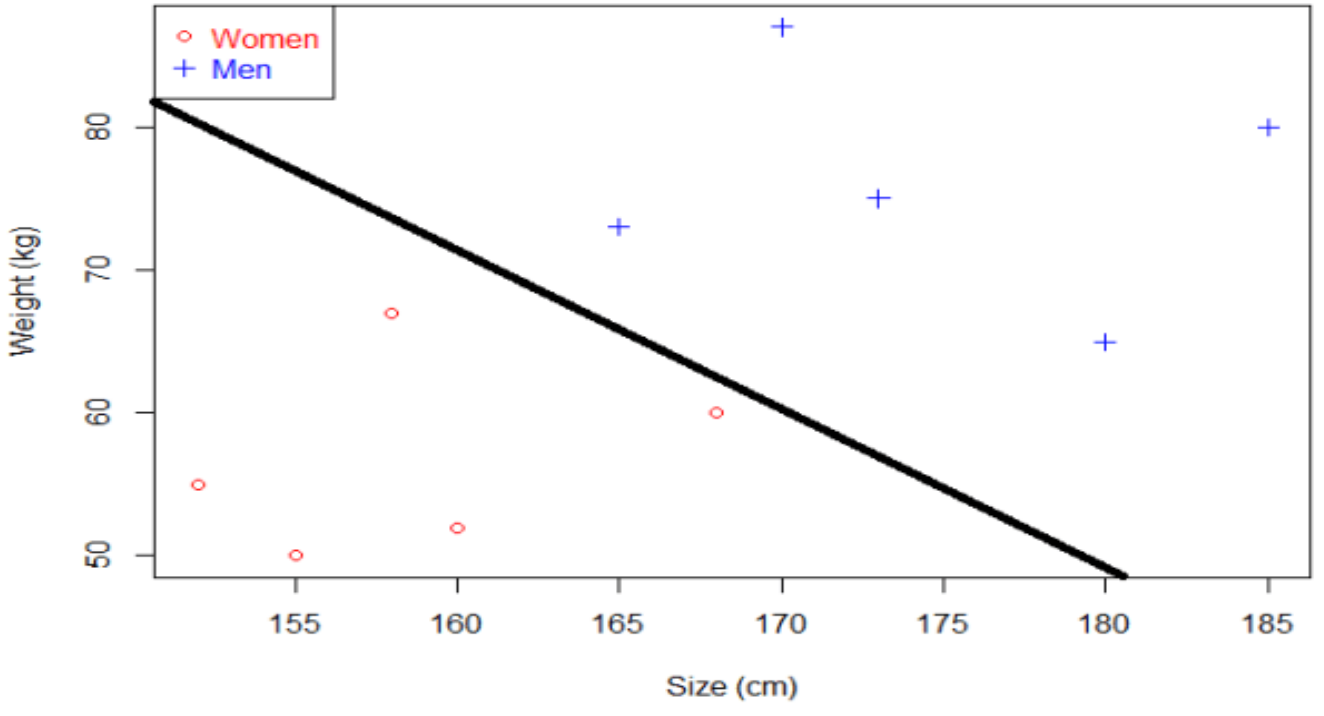


SVM - Understanding the math - Part 1 - The margin

What is a separating hyperplane?

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

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



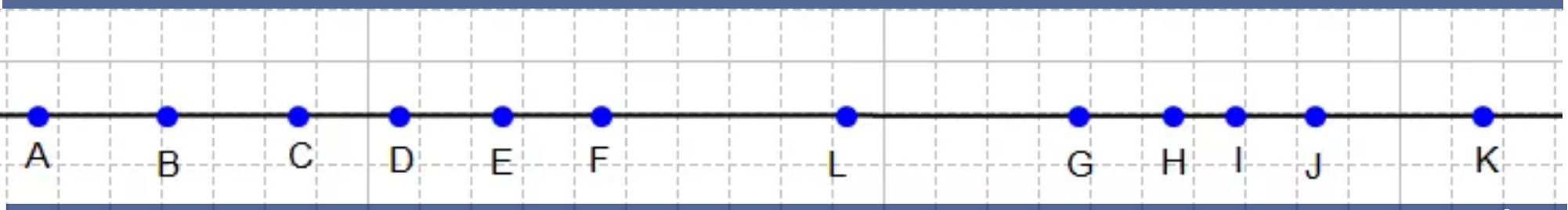
SVM - Understanding the math - Part 1 - The margin

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

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

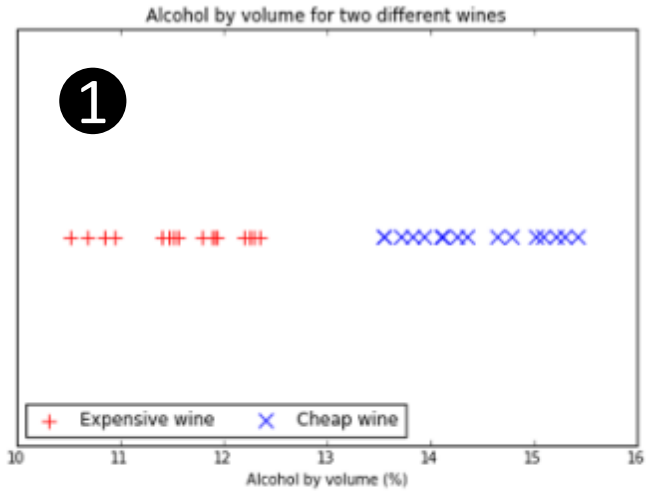
An hyperplane is a generalization of a plane.

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

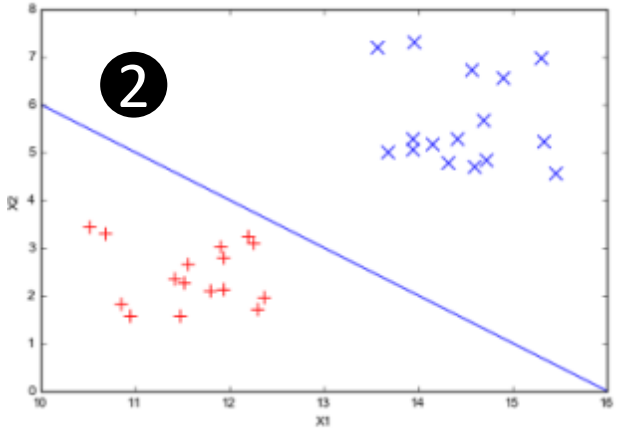


The point L is a separating hyperplane in one dimension

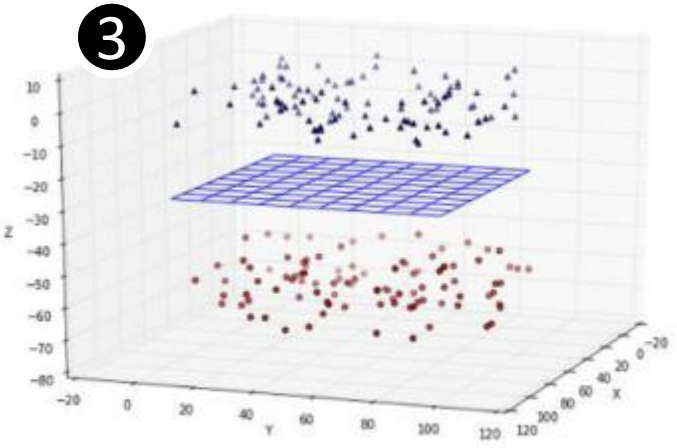
SVM - Understanding the math - Part 1 - The margin



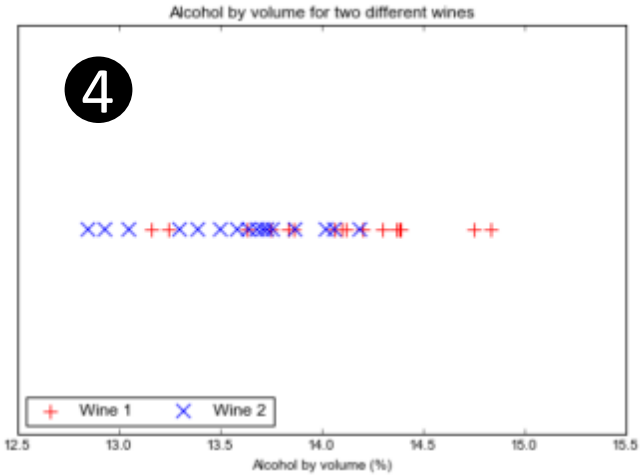
An example of linearly separable data



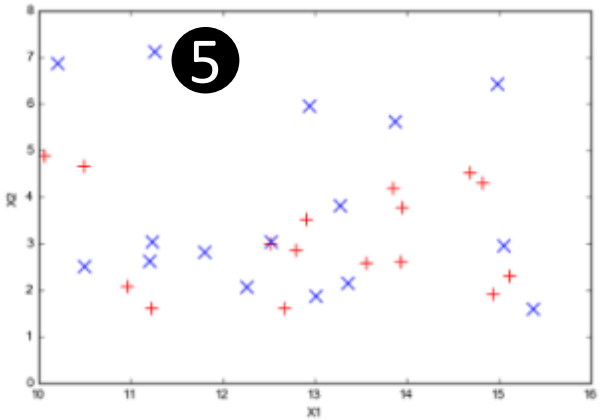
An example of linearly separable data by a line



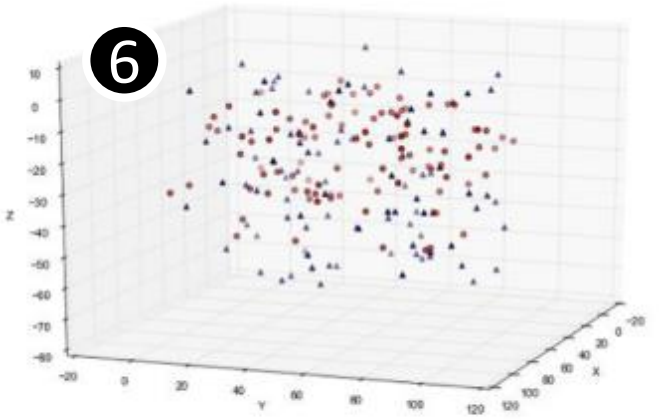
An example of linearly separable data by a plane



An example of non-linearly separable data



An example of non-linearly separable data in 2D

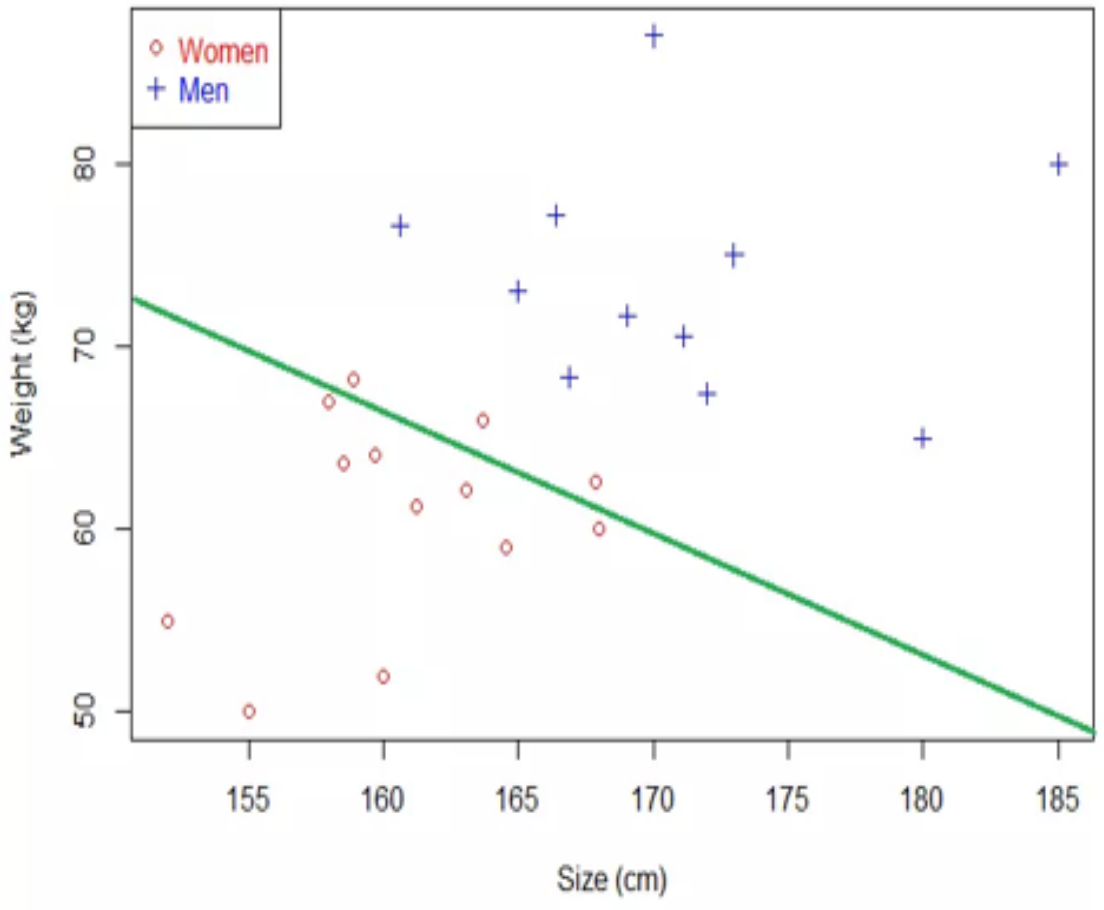


An example of non-linearly separable data in 3D

SVM - Understanding the math - Part 1 - The margin

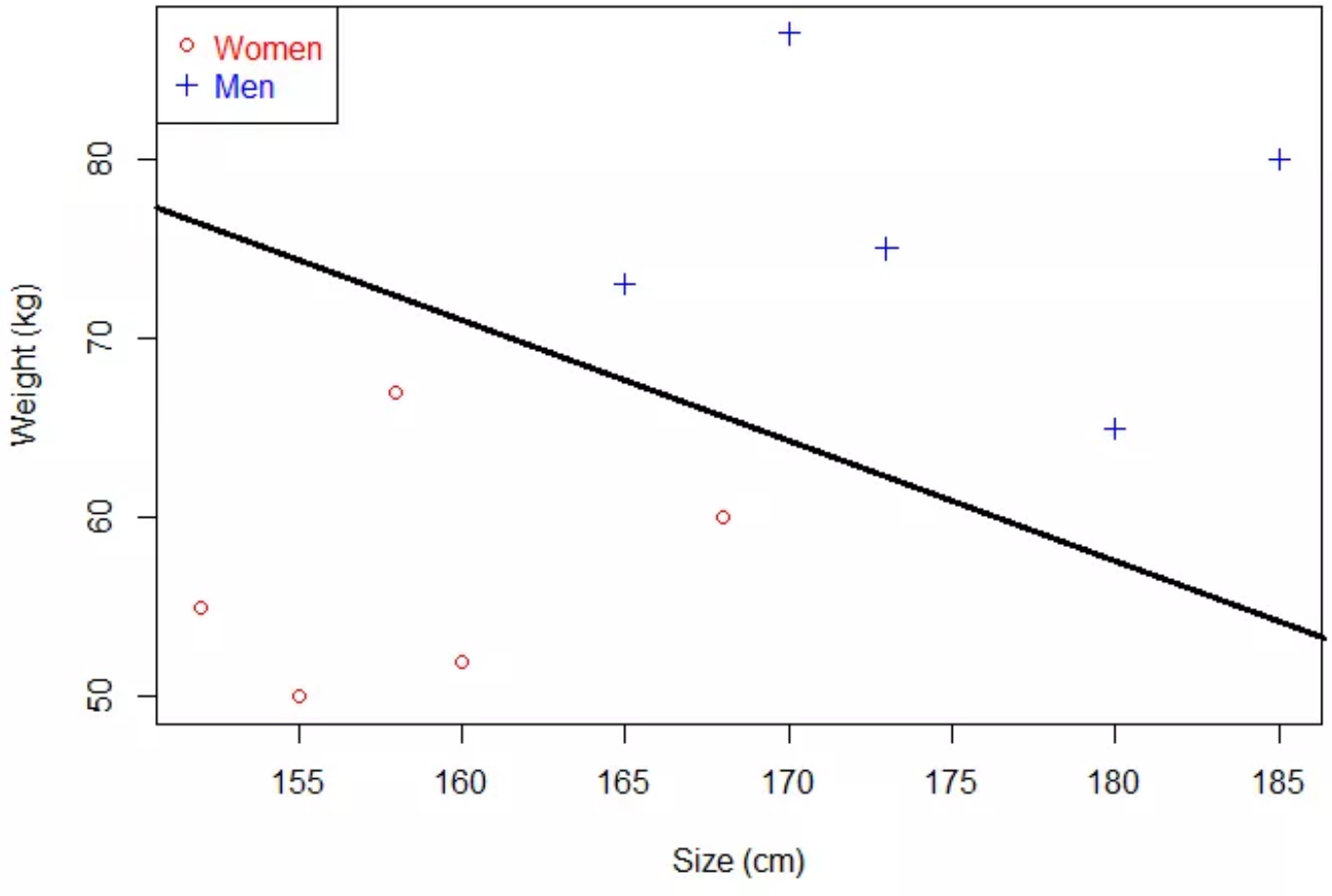
- Suppose we select the green hyperplane and use it to classify on real life data.

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



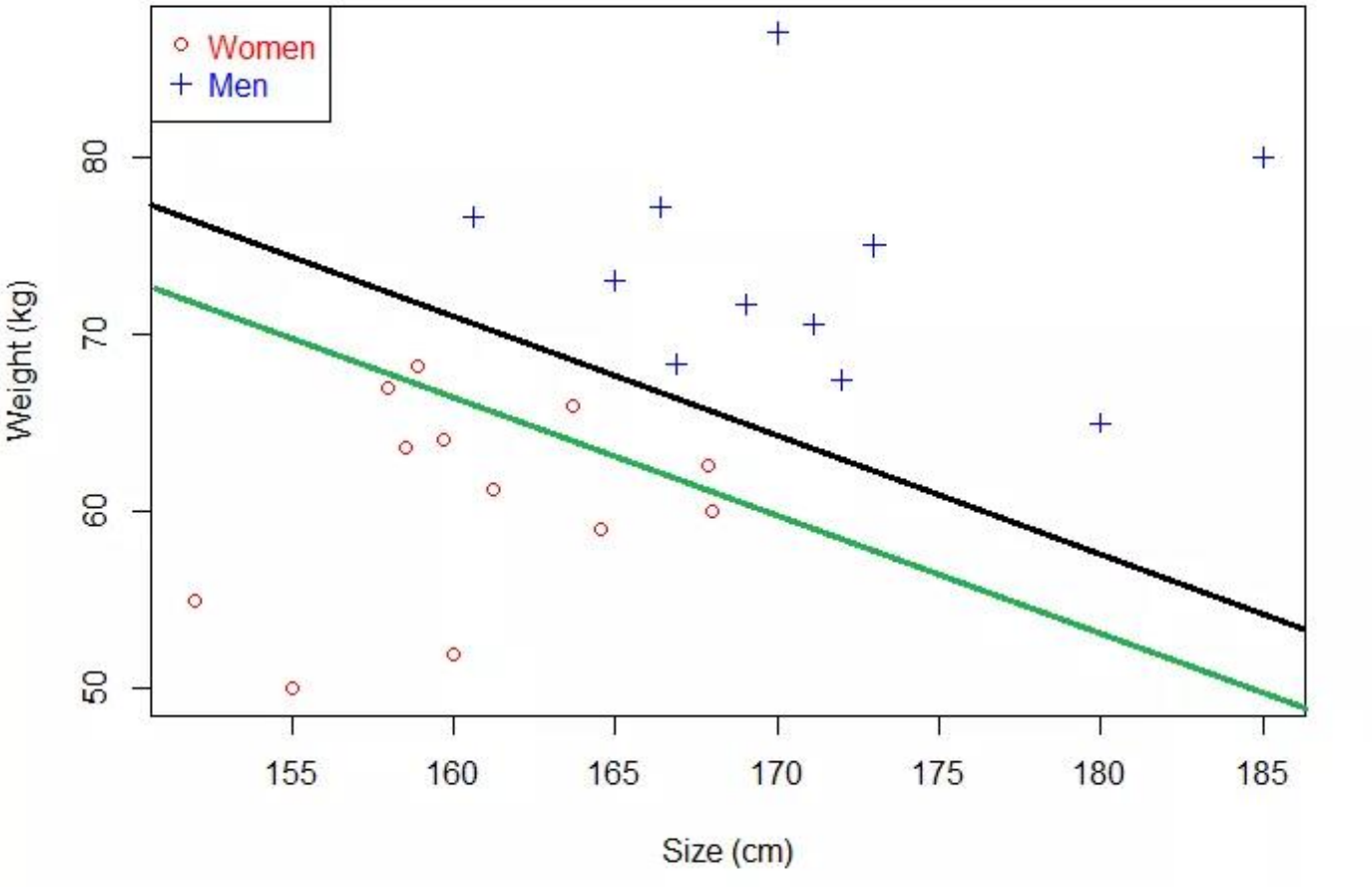
SVM - Understanding the math - Part 1 - The margin

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



SVM - Understanding the math - Part 1 - The margin

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



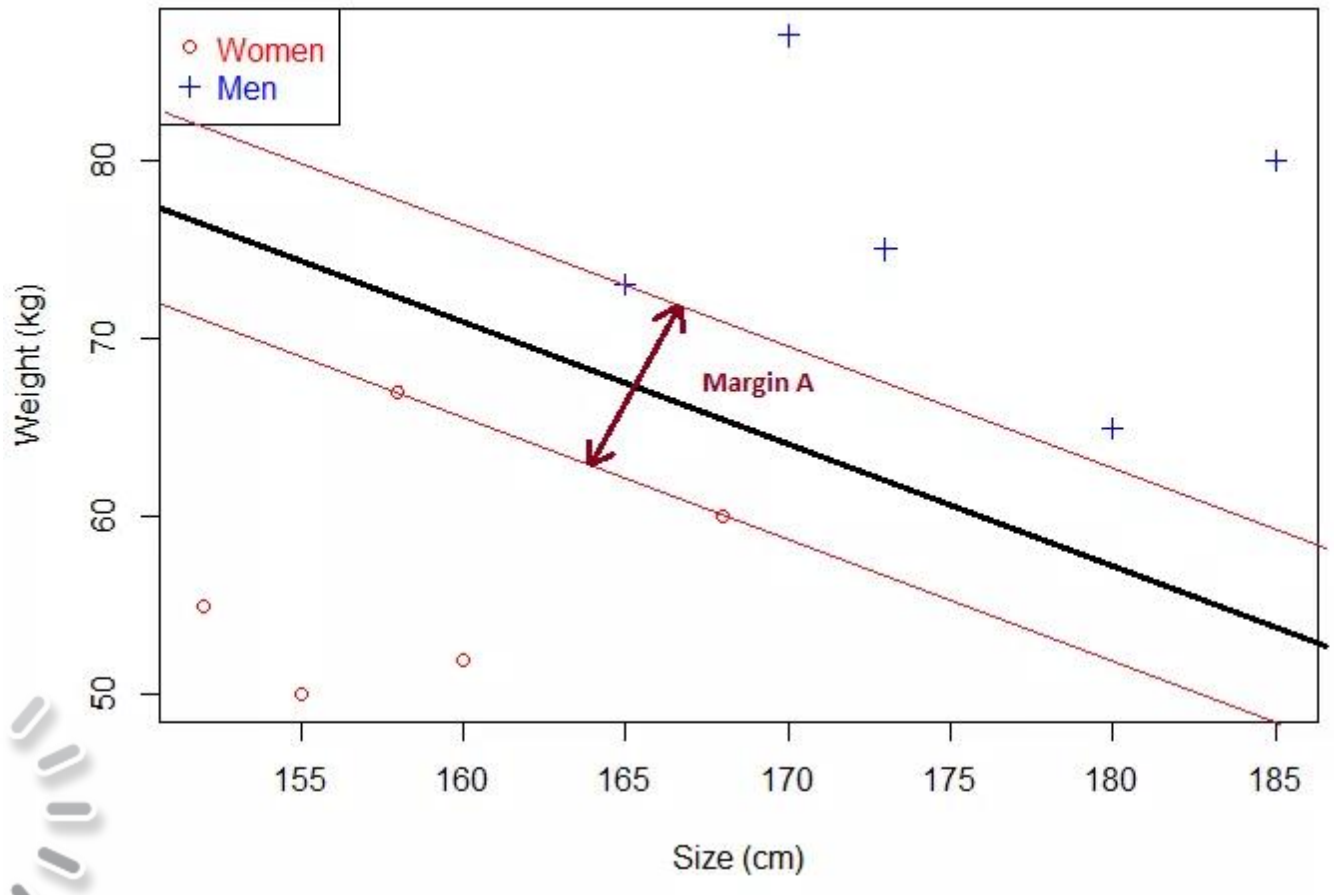
- *The black hyperplane classifies more accurately than the green one*

SVM - Understanding the math - Part 1 - The margin

- That's why the objective of a SVM is to find the optimal separating hyperplane:
    - because it correctly classifies the training data
    - and because it is the one which will generalize better with unseen data

SVM - Understanding the math - Part 1 - The margin

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



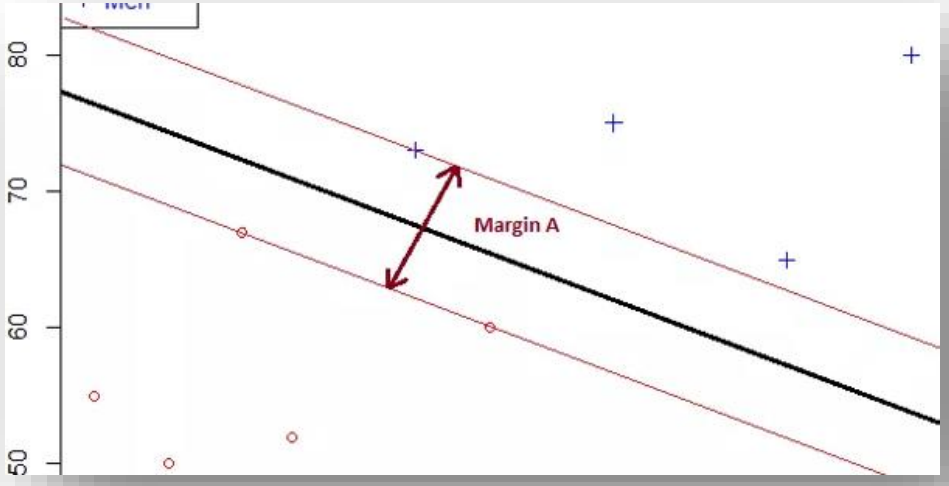
SVM - Understanding the math - Part 1 - The margin

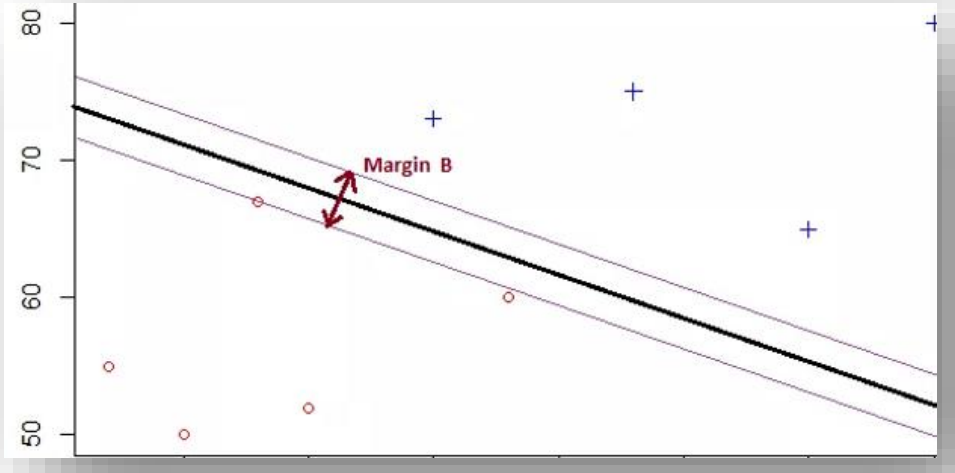
- Given a particular hyperplane, we can compute the distance between the hyperplane and the closest data point. Once we have this value, if we double it we will get what is called the ***margin***.

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

SVM - Understanding the math - Part 1 - The margin

- As you can see, Margin B is smaller than Margin A.
- We can make the following observations:
    - If an hyperplane is very close to a data point, its margin will be small.
    - The further an hyperplane is from a data point, the larger its margin will be.
- This means that ***the optimal hyperplane will be the one with the biggest margin***.
- That is why the objective of the SVM is to ***find the optimal separating hyperplane which maximizes the margin of the training data***.





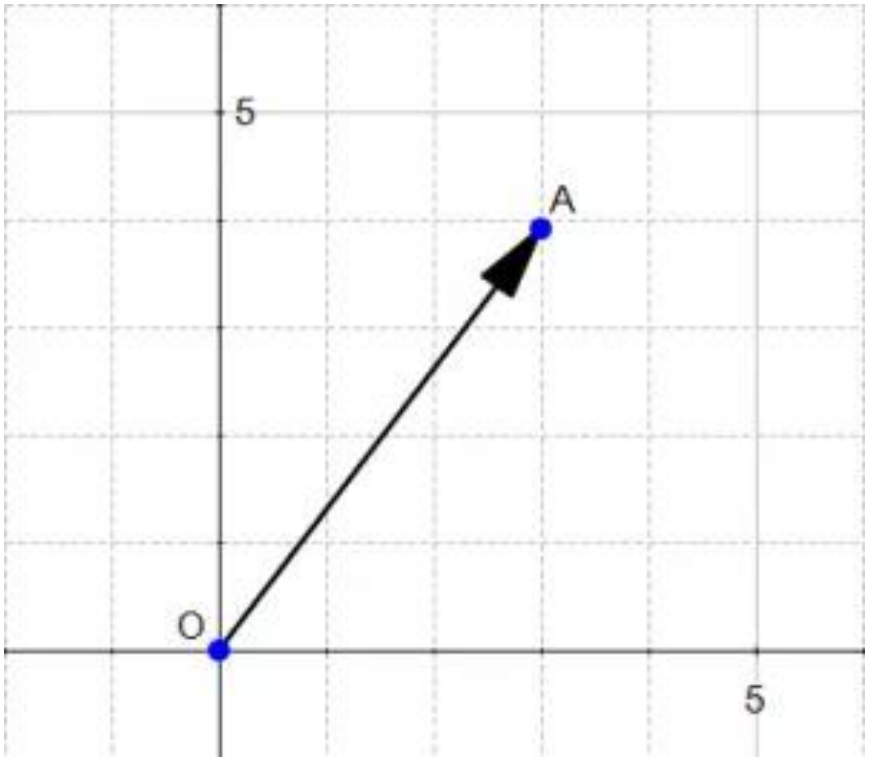
SVM - Understanding the math - Part 1 - The margin

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.

SVM - Understanding the math - Part 1 - The margin

- Definition: Any point $x=(x1,x2), x \neq 0,$ in 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.



SVM - Understanding the math - Part 1 - The margin

- A vector is a mathematical object that can be represented by an arrow.
- In calculations, we denote a vector with the coordinates of its endpoint (the point where the tip of the arrow is).
- If the point at the origin is the point $O(0,0)$ and the point A has the coordinates $(3,4)$ then the vector is the vector $\overrightarrow{OA}$. We can write:

$$\overrightarrow{OA} = (3,4)$$

- We could also give it an arbitrary name such as *a*.

$$a=(3,4)$$

SVM - Understanding the math - Part 1 - The margin

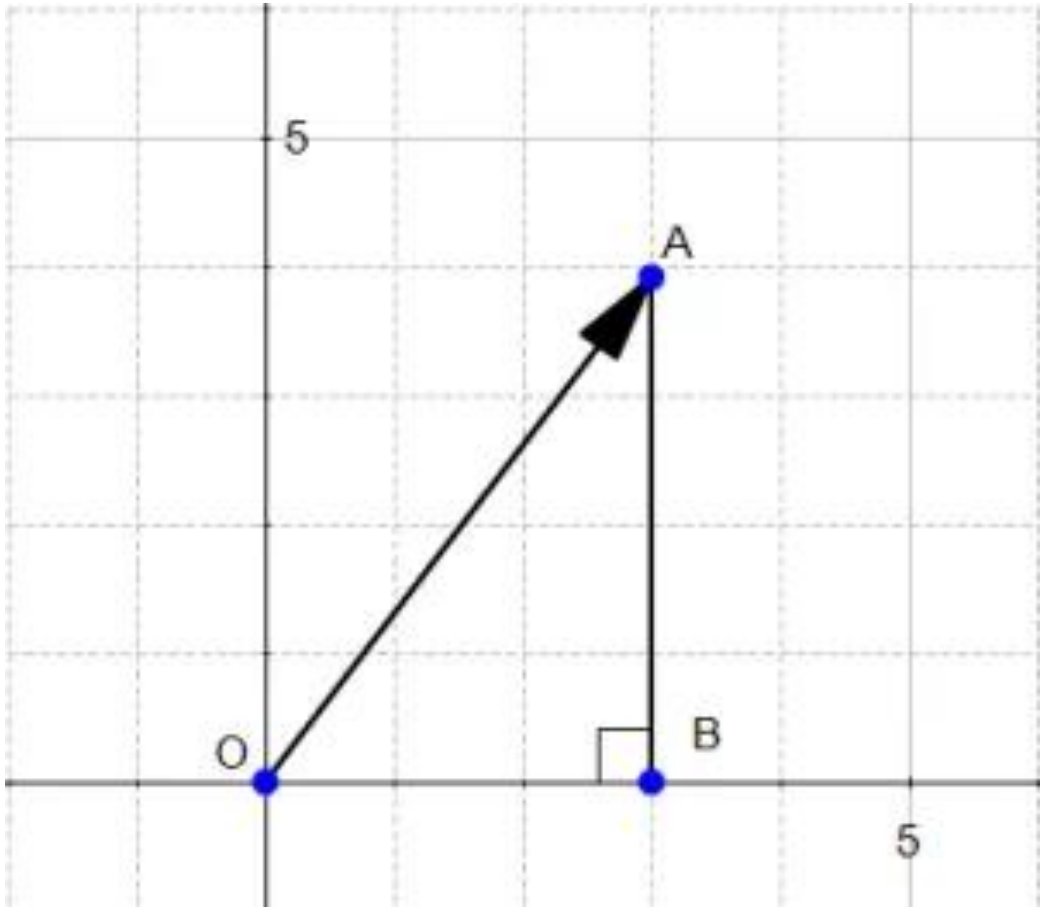
- Definition:

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

SVM - Understanding the math - Part 1 - The margin

1) The magnitude

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



SVM - Understanding the math - Part 1 - The margin

- From Figure 10 we can easily calculate the distance OA using Pythagoras' theorem:

$$\mathrm{OA}^2=\mathrm{OB}^2+\mathrm{AB}^2$$
$$\mathrm{OA}^2=3^2+4^2$$
$$\mathrm{OA}^2=25$$
$$\mathrm{OA}=\sqrt{25}$$
$$\|\mathrm{OA}\|=\mathrm{OA}=5$$

- In general, we compute the norm of a vector $x = (x_1, \ldots, x_n)$ by using the Euclidean norm formula:

$$\|X\| := \sqrt{x_1^2 + \ \ldots + \ x_n^2}$$

SVM - Understanding the math - Part 1 - The margin

2) The direction

- The direction is the second component of a vector. By definition, it is a new vector for which the coordinates are the initial coordinates of our vector divided by its norm.
- Definition :

  The direction of a vector u(u1,u2) is the vector w

$$w = \left(\frac{u_1}{\|u\|}, \frac{u_2}{\|u\|}\right)$$

- Where does the coordinates of w come from ?

SVM - Understanding the math - Part 1 - The margin

- Figure 11 displays the vector $u(u_1,u_2)$ with $u_1=3$ and $u_2=4$
- We could say that :

  Naive definition 1: The direction of the vector 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 :

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

Definition

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

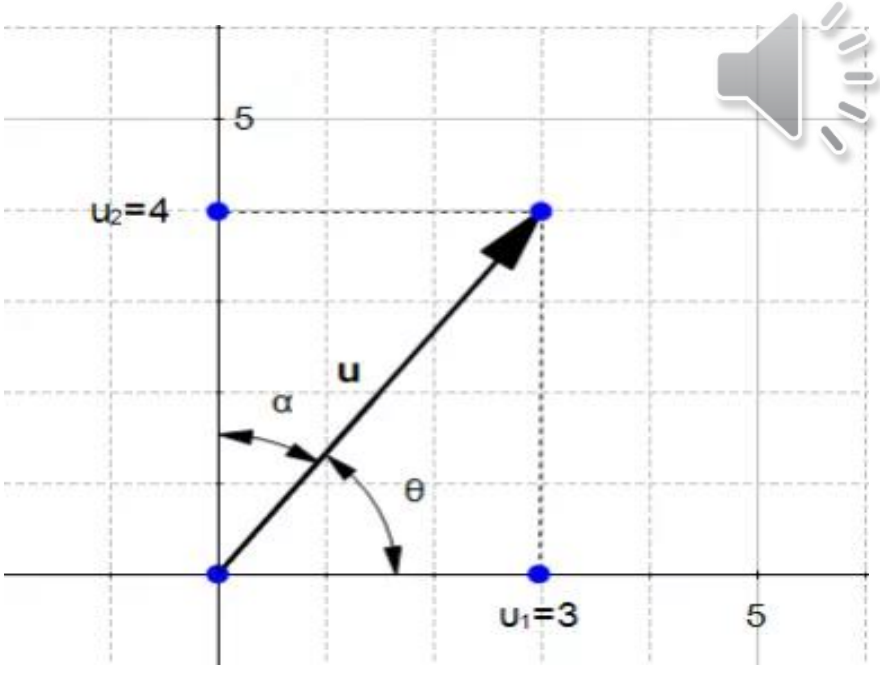


Figure 11-direction of a vector

SVM - Understanding the math - Part 1 - The margin

- In Figure 11 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 u is defined by the cosine of the angle θ and the cosine of the angle α.

SVM - Understanding the math
Part 1 - The margin

- 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.

SVM - Understanding the math - Part 1 - The margin

Computing the direction vector

- We will now compute the direction of the vector u from Figure 11:

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

and

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

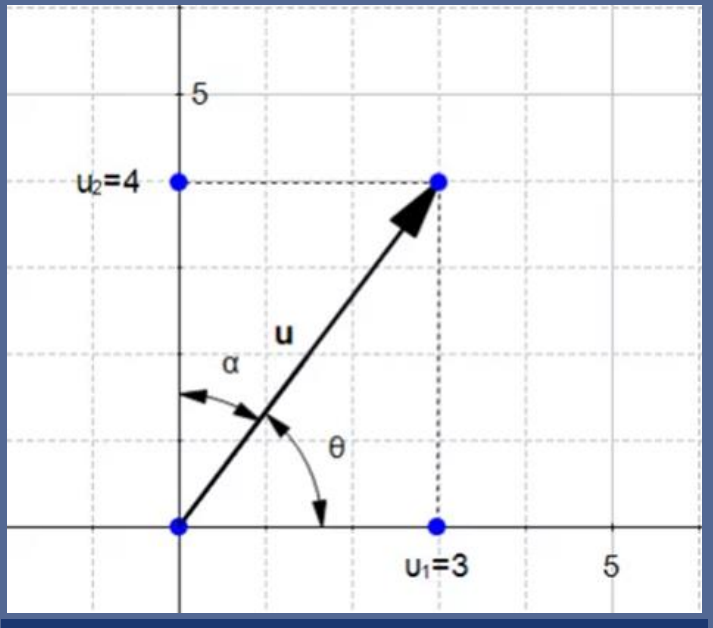


Figure 11-direction of a vector

- The direction of u(3,4) is the vector w(0.6,0.8)

SVM - Understanding the math - Part 1 - The margin

- If we draw this vector we get Figure 12:

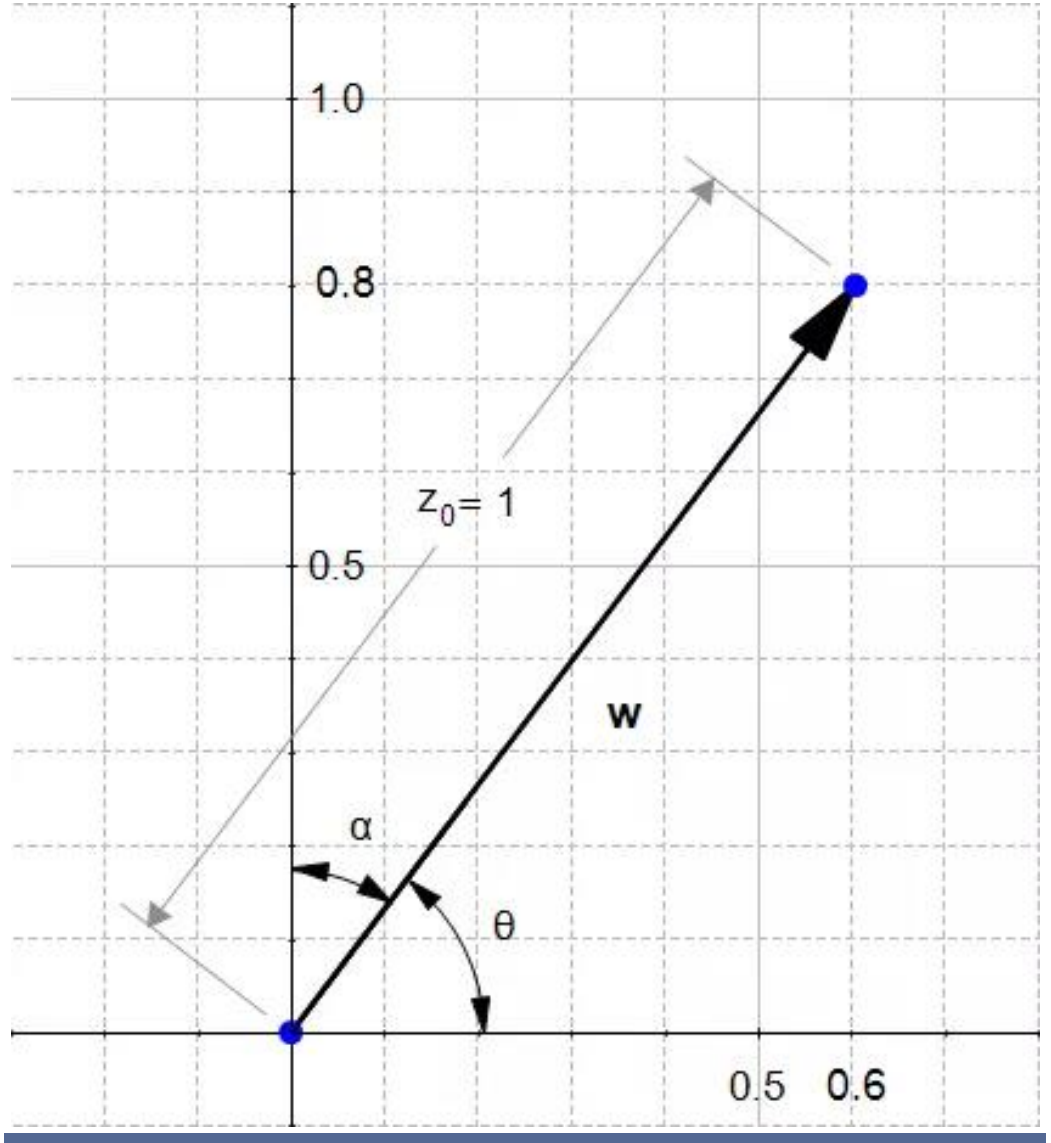


Figure 12

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