"Support - Vector - Machines"

Support vector machines, with respect to supervised learning are able to solve both classification & Regression problems.

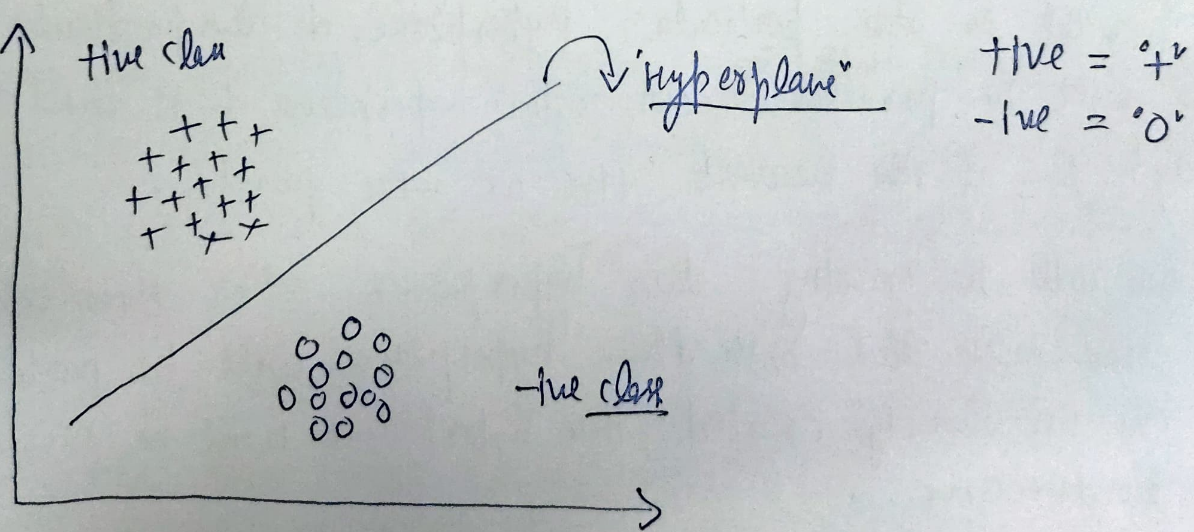
Here we are going to understand.

What are support vectors? What are Hyperplanes?, What is marginal distance?. What are linear separable points? & what are non linear separable points?.

The main aim of Support vector machine is to solve either a "classification problem" or a "regression problem"

Let's have a classification problem with points belonging to two separate classes.

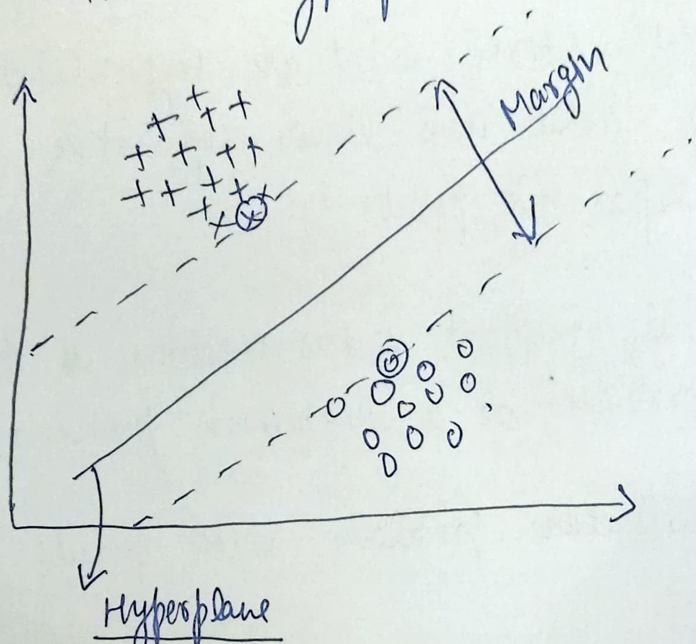
Let's say we have a "+ive class" & a "-ive class"



Now we can classify these two groups of points using a hyperplane. The central line you are seeing over here is a hyperplane. In logistic & linear regression, we have already seen how to create a hyperplane

This is not just the end. SVM makes sure that when we are creating a hyperplane, apart from that, we are also two marginal lines. & these two marginal lines will be having some distance so that both the groups of points are easily linearly separable.

Suppose if we consider a hyperplane as shown below.

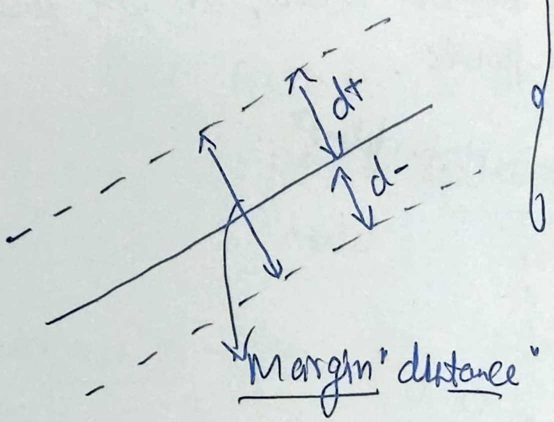


It makes sure that there will be one more plane that will be parallel to this particular hyperplane, & this particular hyperplane will be parallel to our original hyperplane & it will be passing through one of the nearest +ve or -ve points.

So we will be creating two new hyperplanes other than the one we found.. such that both these hyperplanes will be parallel to the one we already created. These hyperplanes will be created in both the directions.

When we are creating these dotted hyperplanes, we will make sure that they pass through one of the nearest +ve & -ve points.

So, this is the basic intuition behind support vector machines, so SVM doesnot focus on creating just a single hyperplane, but instead it also creates two more hyperplanes in a such a way that one of the hyperplane will be passing through the nearest "+ive point" & other hyperplane will be passing through the nearest "-ive point".



When we compute the distance "d+" & "d-" then this whole distance is called "Margin".

Dotted lines we consider as margins & the distance between them is called ["Marginal-distance."].

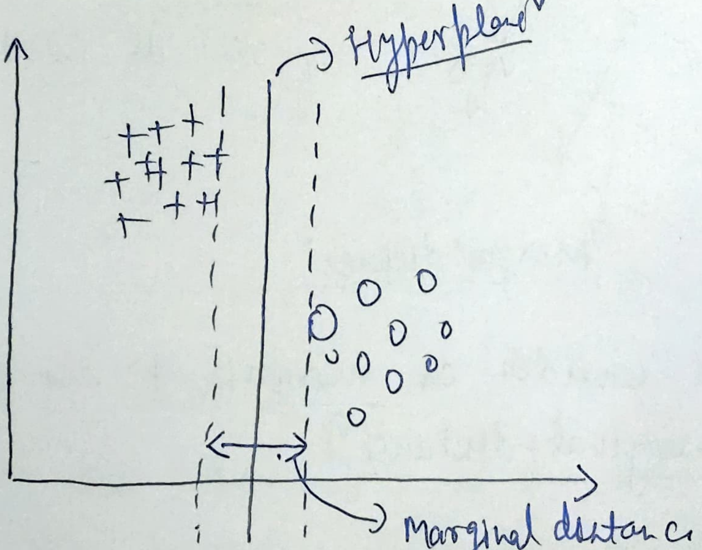
Let's now see what is the significance of this "margin". We always have to create a generalised model. In generalised models we usually get a better accuracy on any kind of data that we use.

Now when we are doing this separation with respect to "+ive" & "-ive". Any point which comes above this hyperplane will be actually classified to [+ive class] & any point which lies below the hyperplane will be classified to the ["-ive class"].

This hyperplane is giving us a cushion to actually divide the points into better clusters

This marginal distance is pretty much important. Now the other question that arises is, apart from this hyperplane, we can also create multiple hyperplanes. But remember that whenever we are creating a hyperplane, we also have to focus on the marginal distance.

Suppose we have considered the same data set & now we have considered a hyperplane as shown below, as we could have multiple hyperplanes to separate the points.

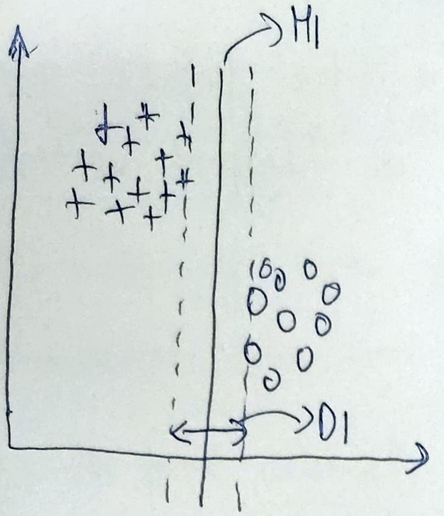


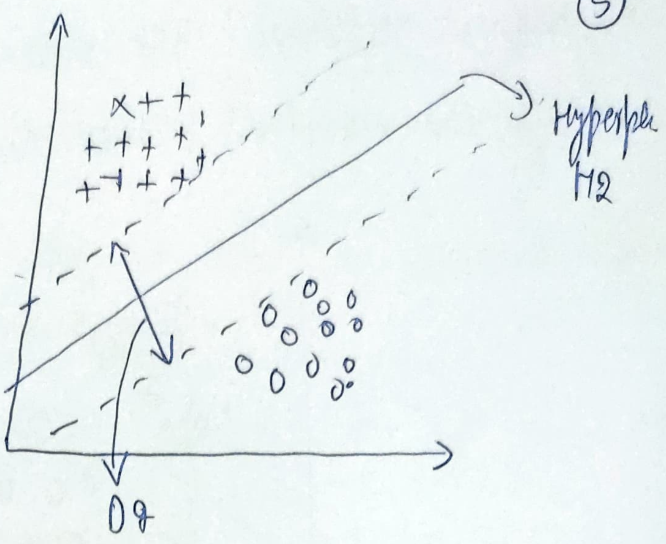
Here you can see that the marginal distance is less than the one we got in the previous example. Our main aim should be that you need to maximize this distance (marginal-distance).

So, based on this, we select the hyperplane which has maximum marginal distance.

Note ⇒ These all techniques are applicable only if the points are linearly separable.

PTO





Since $D_2 > D_1$ so we will select 'H_2'

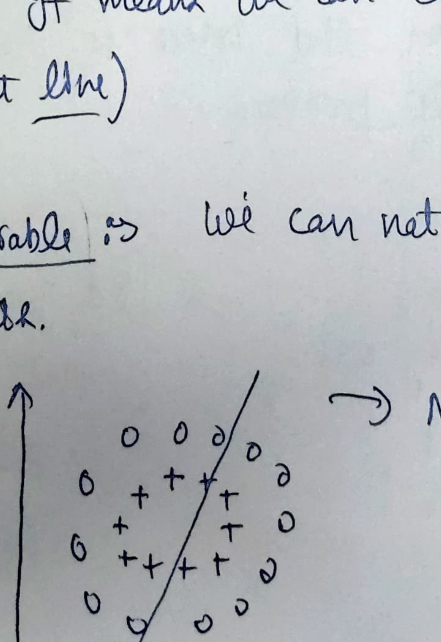
Here in D_2, we will be getting a more generalised model.

"Marginal distance" :- The two parallel planes that we are creating with respect to the most nearest "+ve point" & the most nearest "-ve point", the distance between them is actually the marginal distance.

Linearly separable :- It means we can easily separate the points using a line. (straight line)

Non-linearly separable :- We can not use a straight line to separate the points.

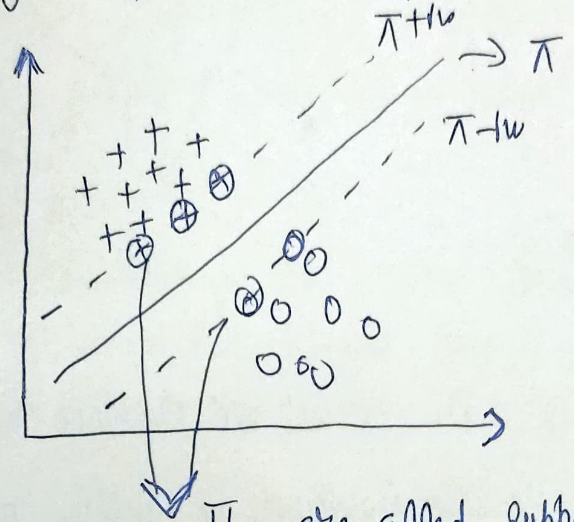
Eg :-



→ Non linearly separable.

We cannot draw a straight line to separate them.

→ Support vectors :→ The nearest +ve & -ve points passing through the marginal planes are known as "support vectors"



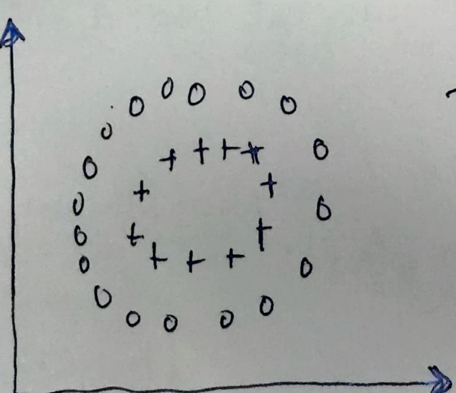
They are called support vectors.

There could be multiple points as well

Support vectors are the points that are passing through the marginal hyperplanes".

Note :→ support vectors help us to determine the marginal distance.

Now another question that arises is, how do we solve non-linearly separable problem.

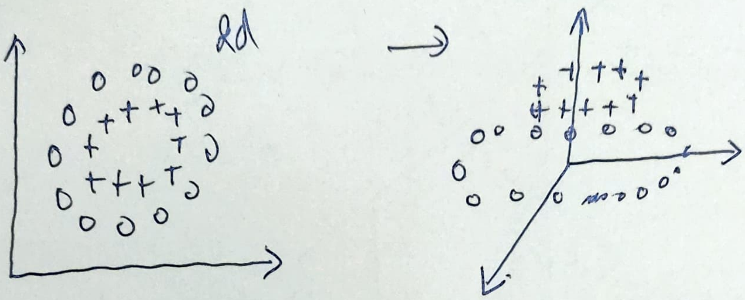


→ [2d graph with respect to +ve & -ve points]

In order to classify non linear datapoints, SVM's are going to use a technique called ["SVM-kernels"].

The main of SVM kernel is that, it tries to convert lower dimensional data into higher-dimensional data

for the same example let's try to convert the "2d" data the '3d' data

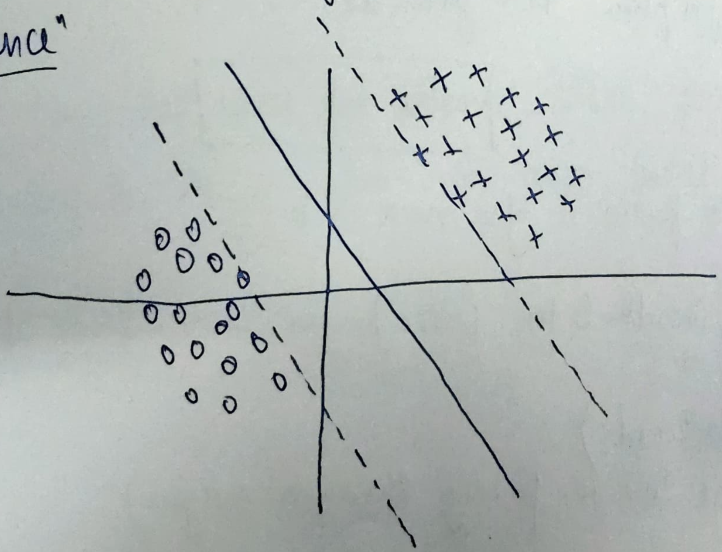


Now we can create hyperplanes to separate them

"This is known as kernel trick"

Mathematical Intuition behind SVM's :->

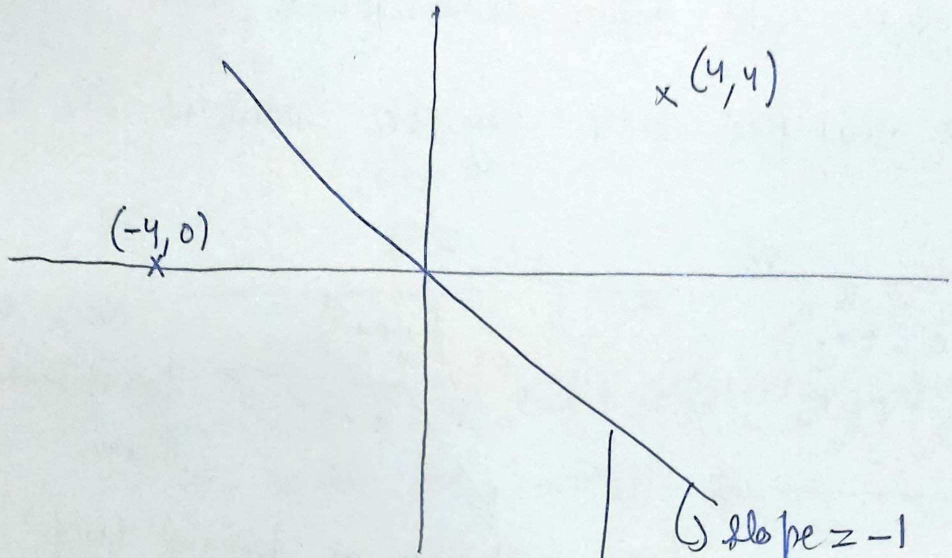
The main aim of "SVM" is to maximize the "marginal distance"



first understand the basic difference between 'logistic-regression' & 'SVM" In both SVM's & 'logistic-regression" we are both finding a straight line to divide the points. But in case.

of SVM, we are adding one more thing the two additional planes that we are creating.

Let's understand the math's behind it :- $\left[\begin{array}{l} y = mx + b \\ y = w^T x + b \end{array}\right]$



Let's say we have two points $(-4,0)$ & $(4,4)$. Let's suppose draw a straight line to divide these two points. Consider the slope to be '-1'.

Equation of a hyperplane is given as:

$$\boxed{w^T x + b = 0}$$

We also know $y = mx + b$

Now for the particular coordinate pair $(-4,0)$ we want to compute the y-value

$m = -1$ (considered)

$C\ b = 0$ (as line is passing through origin)

In short our equation becomes.

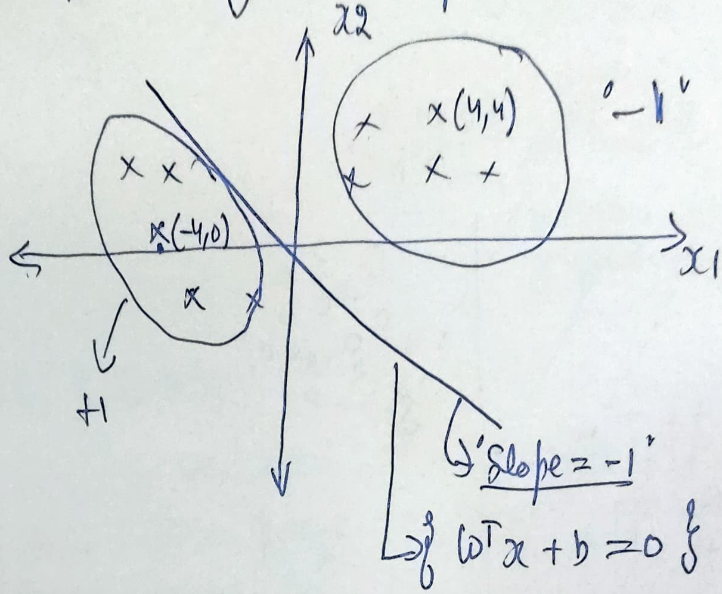
$$y = w^T x$$

Now w.r.t the "w" the 'c' is '0'

$$= \begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -4 & 0 \end{bmatrix}$$

$=$ "y" $\Rightarrow$ a positive value.

Let me come up with any other point over here



& if we try to compute the "y" value, it is always going to be '+ve'

magnitude may change, but sign will always remain the

Now let us go above this ~~point~~ line & let's us try to compute the y value for any point lying above the line.

$$y = w^T x$$

$$\begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} 4 & 4 \end{bmatrix}$$

$$= \text{"}-4\text{"}$$

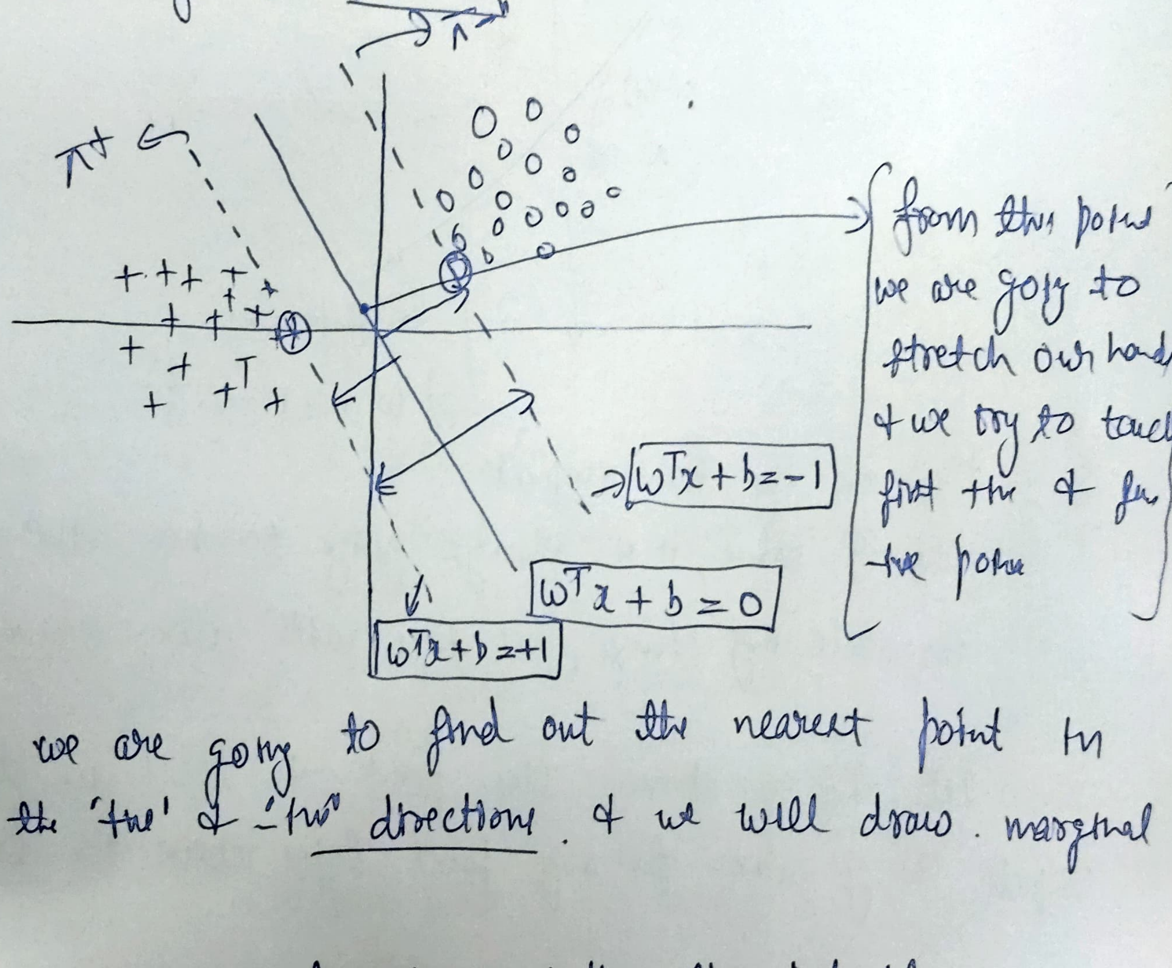
Now if we come up with any point above this line, the 'y' value is always going to be -ve.

So this indicates that the points lying above this line are '-ve' & the points lying below this line are always +ve.

So we can consider these points lying above & below this line as two groups.

Now when I say positive, can I consider it to be "+1" becoz anyhow it is going to be '+ve' & when I say -ve, can I consider it as "-1" becoz still it is going to "-ve" anyways

Now let's go to SVM's

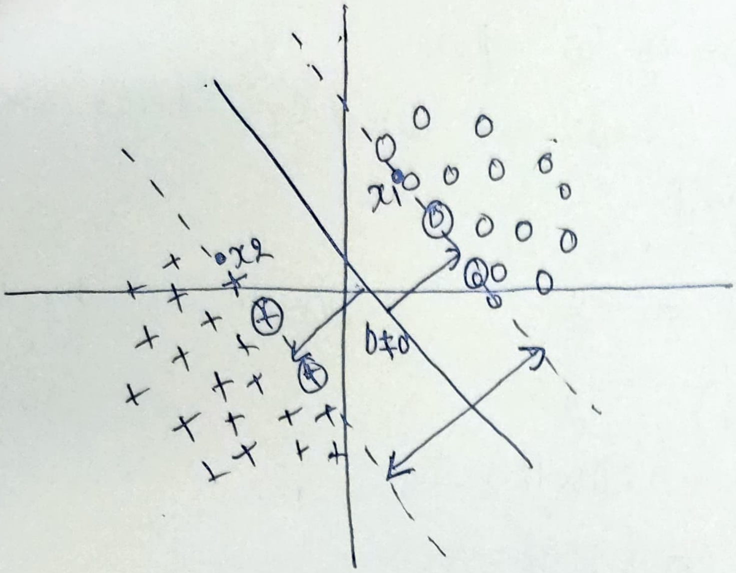


So, we are going to find out the nearest point in both the '+ve' & '-ve' directions. & we will draw marginal plane

In SVM's we are not just computing the hyperplane we are actually computing ["+ve"] & ["-ve"] planes also". to have the maximum distance

Here there will be some 'b' value also.

Now let's compute the ["marginal distance"]



In order to compute the marginal distance we need to basically find the distance between any two corresponding points on these two planes.

Let these points be 'x_1' & 'x_2'

& we need to actually compute the distance between these two points 'x_1' & 'x_2' to find the [marginal distance"]

for 'x_1' we can write the eq as

$$w^T x_1 + b = -1$$

for 'x_2' we can write the eq. as

$$w^T x_2 + b = +1$$

Now we need to compute the difference between x_1 & x_2 to find the marginal distance.

$$\begin{array}{r} w^T x_1 + b = -1 \\ -\; w^T x_2 \pm b = \mp 1 \\ \hline w^T (x_2 - x_1) = 2. \end{array}$$

Now we have only w^T here & we need to remove it

we can remove it by norm of 'w' $\|w\|$
so we cannot remove w^T directly becoz there is some direction involved.

$$w^T(x_2 - x_1) = 2 \quad \text{Dividing b/s by } \|w\|$$

$$\Rightarrow \frac{w^T}{\|w\|}(x_2 - x_1) = \frac{2}{\|w\|}$$

$$\Rightarrow (x_2 - x_1) = \boxed{\frac{2}{\|w\|}}$$

That is the optimization function & we need to maximize it.

So, we need find 'w & b' in such a way that $\frac{2}{\|w\|}$ will be maximized.

$$\boxed{\underset{(w^*,\, b^*)}{\operatorname{argmax}} \frac{2}{\|w\|} \quad \text{such that} \quad y = \begin{cases} +1 & w^Tx + b \geq 1 \\ -1 & w^Tx + b \leq -1 \end{cases}}$$

optimization function. that we want to maximize.