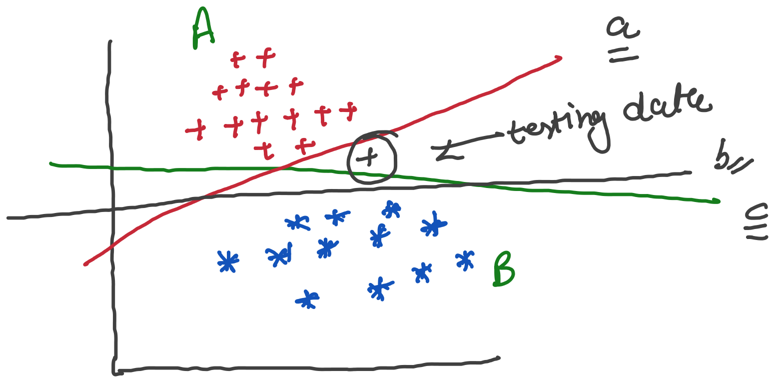
SVM Notes

29 April 2022 13:40

why need of S.V.M. :

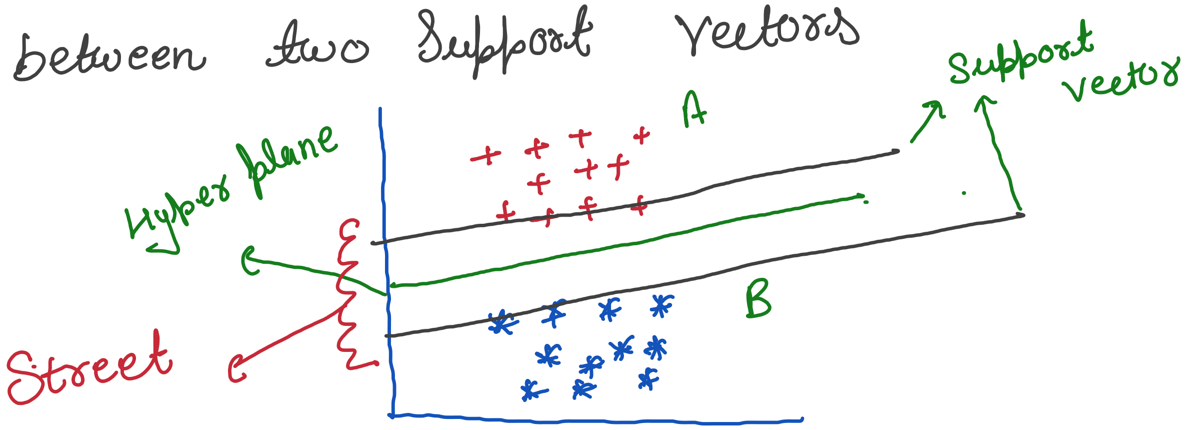
Let us look at logistic regression modes :



We can see that we can get any of these three lines whichever have good accuracy on training data.

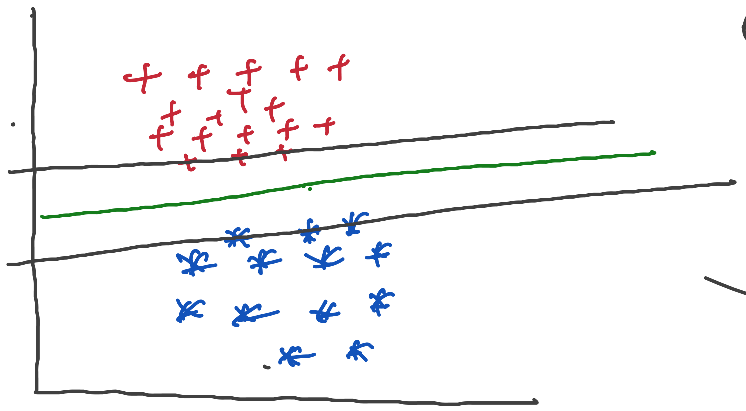
But now look at line a an element comes of class A but model predicted as B.
So, it is giving less accuracy or high error on testing data.

So, SVM solve this problem by finding mid line between two support vectors



Now, let us know about hard Margin and Soft Margin.

HARD MARGIN

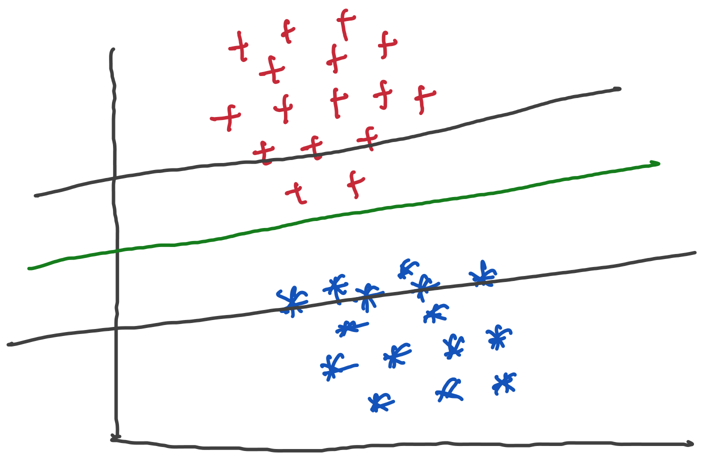


Condition of Hard Margin:

i) Max width

ii) No error on training data

Soft Margin



Condition for Soft Margin:

i) Max Street width

ii) Some error (minimum)

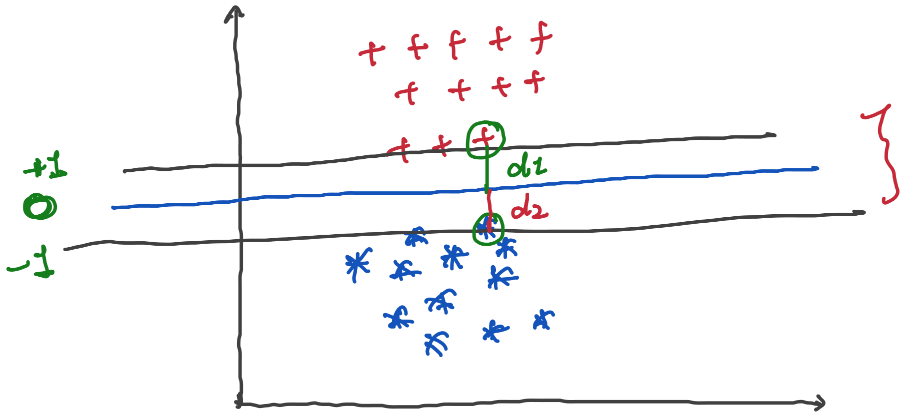
In both of above cases we have to maximize width of Street.

Let do that :

To find width of street we have to find d_1 and d_2.

d_1: distance of point x_1 from plane.

d_2= distance of point x_2 from plane



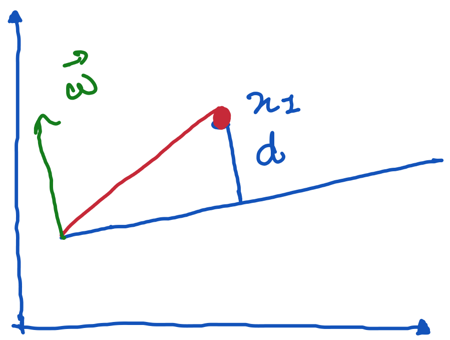
$\}\rightarrow$ Total street width

$= d_1 + d_2$

but d_2 is below so it is negative

The formula $= d_1 - d_2$

$d_1 =$



Projection of x_1 on $\vec{\omega} = d$

$$d = \frac{x_1 \cdot \vec{\omega}}{|\omega|}$$

$$= x \cdot \vec{\omega} \qquad |\omega| = 1$$

assumed

$$d_1 = \omega^T x_1 + b$$

$$d_2 = \omega^T x_2 + b$$

Width of street :

$$\omega^T x_1 + \cancel{b} = 1 \rightarrow \text{assumed } 1$$

$$\omega^T x_2 + \cancel{b} = -1 \rightarrow \text{assumed } -1$$

$\}$ when hyperplane at 0

$$- \qquad - \qquad +$$

$$\omega^T (x_1 - x_2) = 2$$

$$\boxed{x_1 - x_2 = \frac{2}{\omega} = \frac{2}{|\omega|}}$$

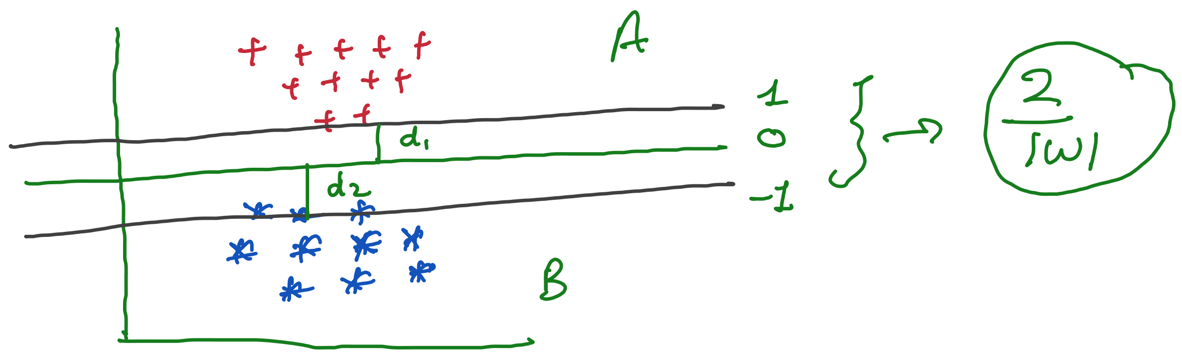
($\because$ street of width can't be negative)

Our work is to maximize it

So :- $\max\left(\frac{2}{|w|}\right)$.

Now Let us find Loss and error:

we got this diagram and street width



for class A : $\boxed{y=1}$

Correct output is : $w^T x_1 + b \geq 1$

error : $w^T x_1 + b < 1$

for class B : $\boxed{y=-1}$

Correct output is : $w^T x_2 + b \leq -1$

Error : $w^T x_2 + b > -1$

Now Let us find generalized form by multiplying y_i.

when $y_i = 1$ The correct is :

$$y_i * w^T x + b \geq 1 * y$$

$$y_i (w^T x + b) \geq 1 * 1$$

$$y_i (w^T x + b) \geq 1 \longleftarrow (1)$$

When $y_i = -1$

$$y_i * (w^T x + b) \leq -1 * y_i$$

$$y_i (w^T x + b) \geq -1 * -1 \quad (\because \text{sign change})$$

$$y_i (w^T x + b) \geq 1 \quad \text{—— (2)}$$

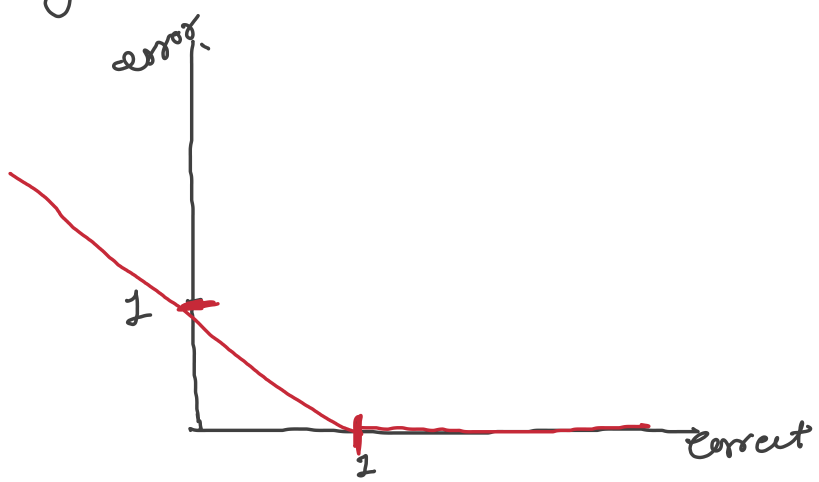
eqn (1) and (2) are same So generalize form is

$$y_i (w^T x + b) \geq 1 \quad \text{for correct}$$

Similarly for error:

$$y_i (w^T x + b) < 1$$

Now from Generalize eqn we can plot Graph



This is called hing loss

Now let us look at error.

So :→

$$\text{error} = \begin{cases} Z & y_i (w^T x + b) < 1 \\ 0 & y_i (w^T x + b) \geq 1 \end{cases}$$

Now this is for single point let us calculate for all m points.

$$\text{error} = \frac{1}{m}\sum_{i=1}^{m} z_i$$

we have to minimize as discussed first

$$\Rightarrow \quad \min\left(\frac{1}{m}\sum_{i=1}^{m} z_i\right)$$

Let us see out cost fn

$$= \max\left(\frac{2}{|w|}\right) + \min\left(\frac{1}{m}\sum_{i=1}^{m} z_i\right)$$

Now to give priority to error minimization we add a hyperparameter c.

$$= \max\left(\frac{2}{|w|}\right) + c \, \min\left(\frac{1}{m}\sum_{i=1}^{m} z_i\right)$$

c value is large:

Priority to minimization of error is high so error will be minimized fcausly.

It goes towards to 0.

which we know if error 0 then it is

Hard Margin.

C is low :

Low priority to minimization of error

So, some error can come means Soft Margin.

Now we can regulate Soft Margin and Hard margin using <u>C</u>.