Support Vector Machines

SVM → 90's - 1995

↑

Small datasets

applicable
- → classification → binary / multiclass
- → regression
- → imag processing → SVM

Linear
Non-linear

Plan Of Attack

1) Maximal Margin Classifier → SVM → Hard Margin SVM

improve ↓

2) Soft margin SVM → Support Vector Classifier → SVC → linear kernel

improve ↓

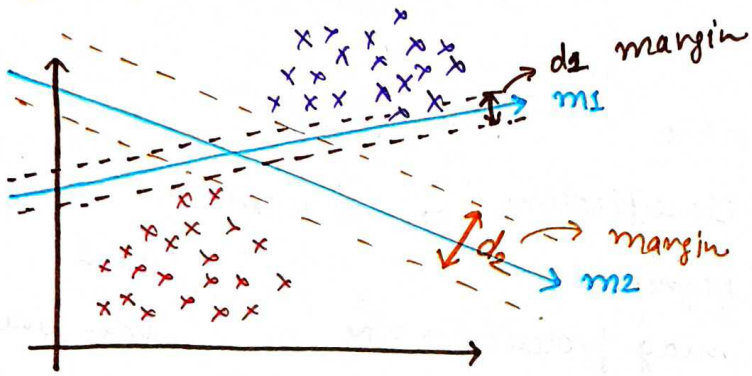
3) SVM → Kernels → non-linear

4) SVM for multiclass setup

5) SVR

Maximal Margin Classifier

cgpa | iq | Placement



d_2 margin > d_1 margin

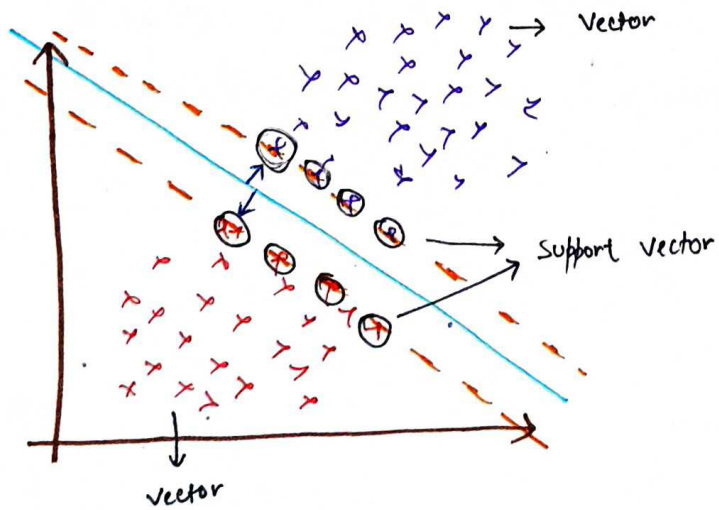
m1 or m2 ?

m2 line is best than m1 line because m1 line is more closure to data and create problem in test data or unknown data

basic requirement

* Linearly seprable

Support Vector



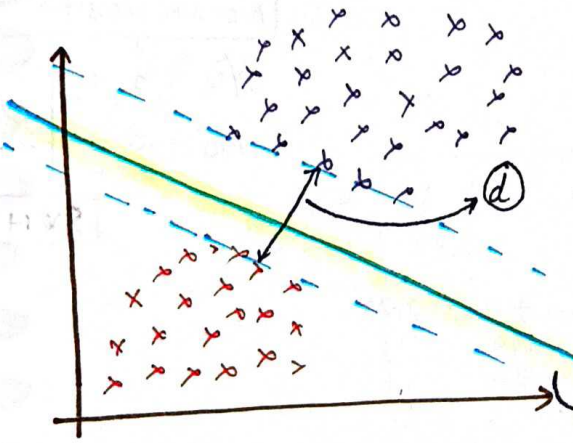
cgpa | iq | placement

Vector ← [- -] -

- - -

Support Vectors: Support Vectors are those point where we can find first points when moving toward direction of point. and those points are end of margin.

Mathematical Formulation

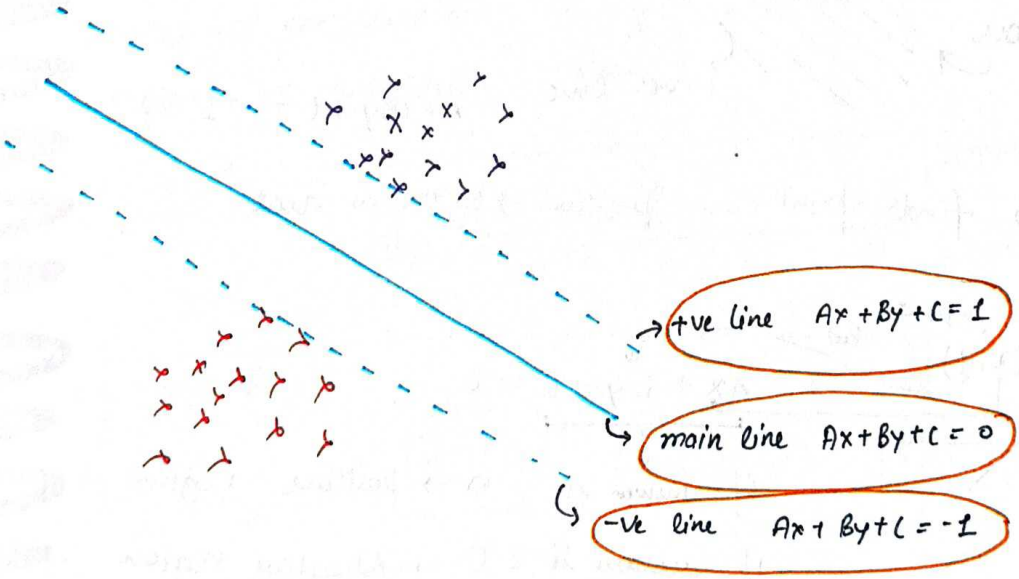


We want a line which have max margin (d) and follow given condition.

given condition: points not cross margin line.

main model: predict output

$Ax + By + C = 0$



* Why we take $Ax + By + C = (1)$ and $Ax + By + C = (-1)$.

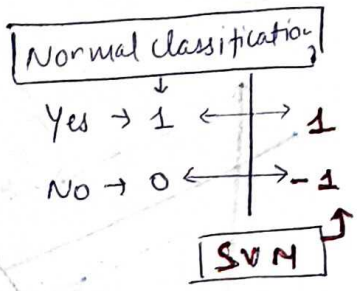
?

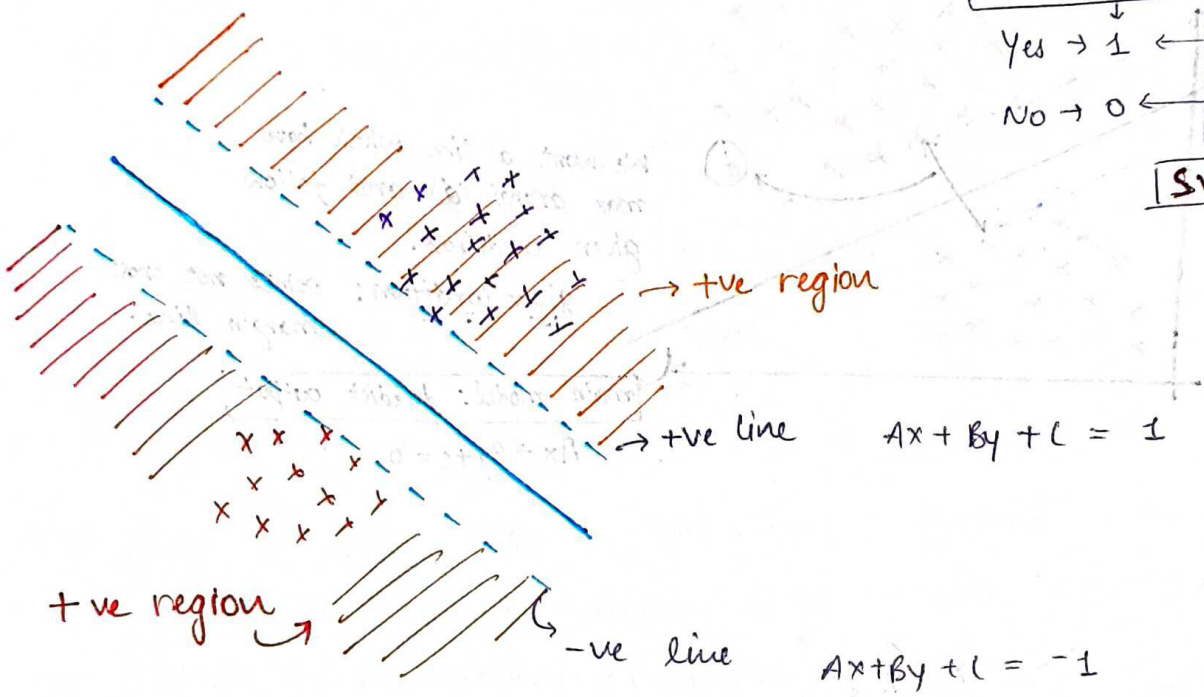
- Because for main line will be in centre.

* Why (1 and -1) why not (2 and -2) or (20 and -20)?

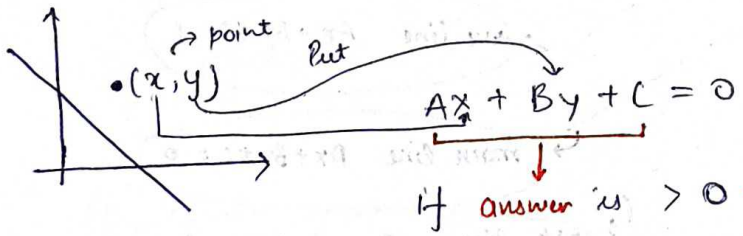
- it's doesn't matter, output will be same.

Condition : Data points not cross margin lines.





* How to find point in positive region or not?



if answer is > 0 → positive region

if answer is < 0 → Negative region

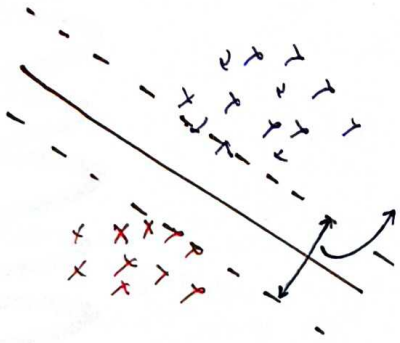
condition 1

$Ax + By + C \geq 1$ → Point in Positive Region or on margin line.

condition 2

$Ax + By + C \leq -1$ → Point in Negative Region or on margin line.

Summary



avg max d

such have

ABC

↑

want → follow condition

if

$y_i = 1 \quad Ax_{1i} + Ax_{2i} + C \geq 1$

$y_i = -1 \quad Ax_{1i} + Ax_{2i} + C \leq -1$

* Single eqn or condition to find where point is in positive region or negative region.

$$y_i (Ax_{i1} + B_{i2} + C) \geq 1$$

→ +ve

→ -ve

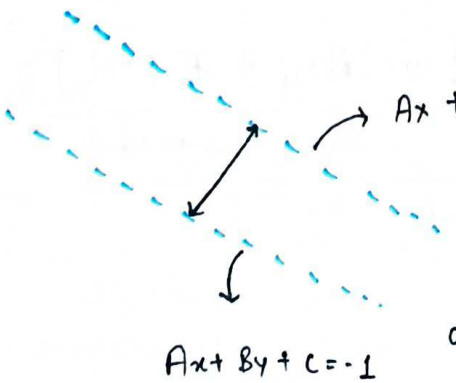
arg max d

A, B, C

such that

$y_i (Ax_{i1} + Bx_{i2} + C) \geq 1$ → Solved condition part

Find distance between two margin line.



$Ax + By + C = 1 \Rightarrow Ax + By + C - 1 = 0$

$ax + by + c_2 = 0$

$ax + by + c_1 = 0$

} Formula of line for distance btn 2 parallel line.

$Ax + By + C = -1$

$Ax + By + C + 1 = 0$

* distance between two parallel line

$$d = \frac{|C_1 - C_2|}{\sqrt{a^2 + b^2}}$$

* for parallel margin line

$$d = \frac{|C+1 - (C+1)|}{\sqrt{A^2 + B^2}} = \frac{2}{\sqrt{A^2+B^2}} = d$$

Final Formula

$$\underset{ABC}{\text{argmax}} \quad \frac{2}{\sqrt{A^2+B^2}} \quad \text{given} \quad \left\{ y_i (Ax_{1i} + Bx_{2i} + C) \geq 1 \right\}$$

How to Solve this ?

constrained optimization problem

$$\underset{ABC}{\text{arg max}} \quad \frac{2}{\sqrt{A^2+B^2}} \quad \text{given} \quad \underbrace{\left\{ y_i (Ax_{1i} + B_{2i} + C) \geq 1 \right\}}_{\text{linear constrained}}$$

quadratic (raise to the power 2)

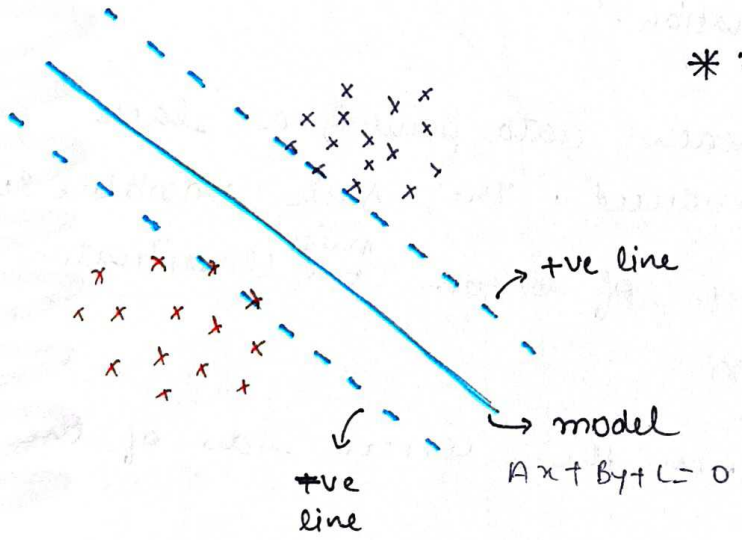
quadratic programming for loss-function.

$$L = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)$$

$$L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log \hat{y}_i + (1-\hat{y}_i)(\log(1-y_i))$$

use gradient descent in LR and Logistic Reg.

Prediction



* margin only use in training not in test (Prediction)

(8, 80)

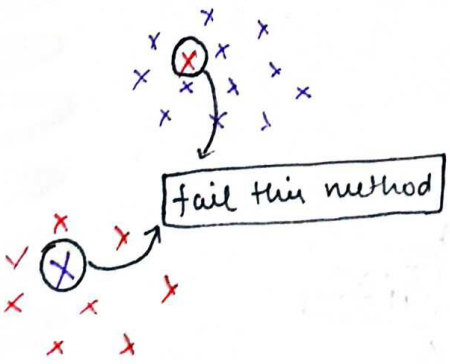
$Ax8 + Bx80 + C \geq 0$ → Placement

≤ 0 Not placement

Why this method called Hard margin SVM?

→ Because its seprate 2 category Hardly and point doesn't cross margin line.

Problem with Hard Margin SVM

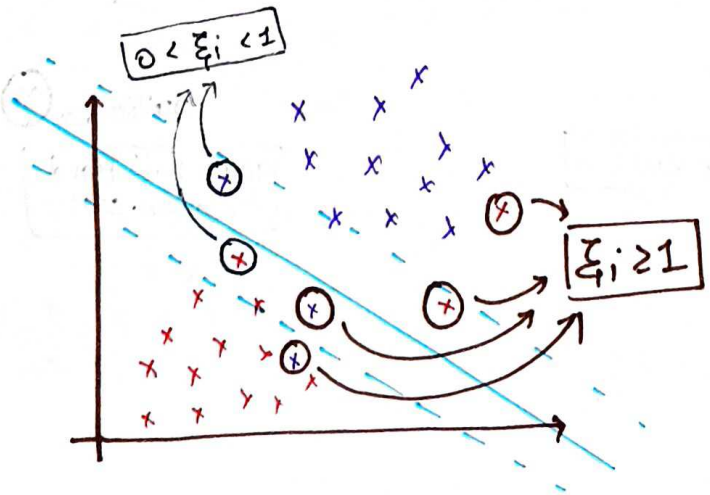


Slack Variable → ξ_i → Misclassification score
↳ Hinge Loss

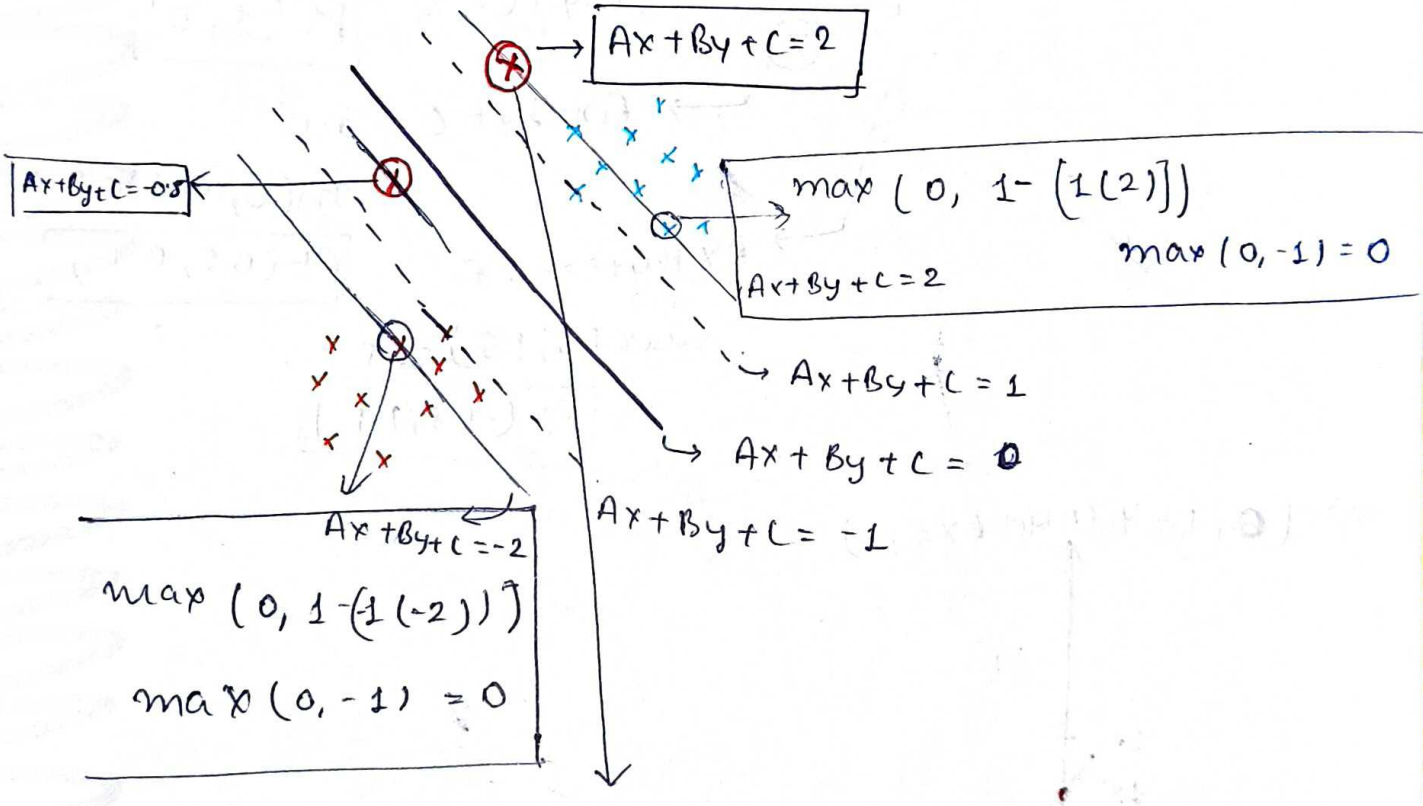
The concept of slack variable was introduced by Vladimir Vapnik in 1995 and is used in the formulation of the "soft-margin" SVM to handle cases where data is not linearly separable, or when one allows for some degree of error in classification.

Mathematically, for each data point i, a slack variable $\xi_i \geq 0$ is introduced. The slack variable ξ_i measures the degree of ~~error~~ miss classification of the data point x_i.

- $\xi_i = 0$ if x_i is on the correct side of the margin
- $0 < \xi_i < 1$ if x_i on the correct side of the hyperplane but on the wrong side of the margin
- $\xi_i \geq 1$ if x_i is on the wrong side of the hyperplane i.e it is misclassified.

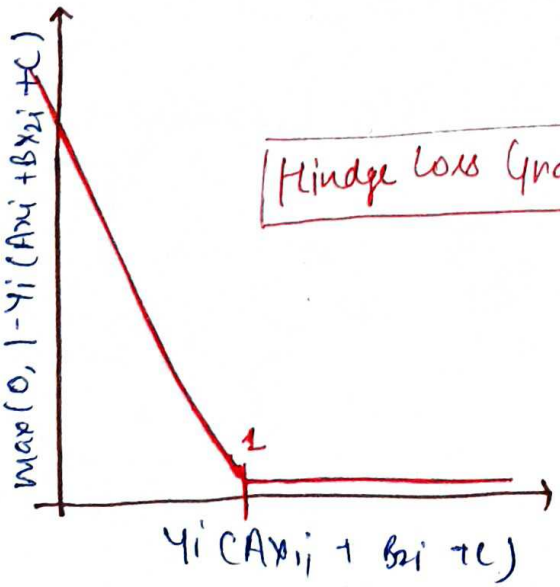


Hinge Loss / Slack $= \max(0, 1 - y_i(Ax_{1i} + Bx_{2i} + C))$

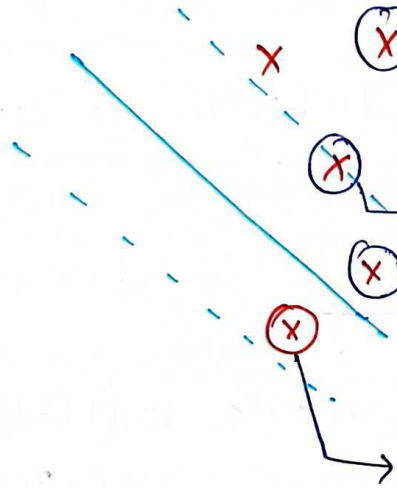


$\max(0, 1 - ((-1)(2)))$

$\max(0, 3) = 3$ ⇒ miss classify



Hinge Loss Graph = ξ_i



$Ax+by+c=2$ $\quad m(0,-1)=0$

pt (2, 0)

$Ax+by+c=1$ pt (1, 0)

$Ax+by+c=0.5$

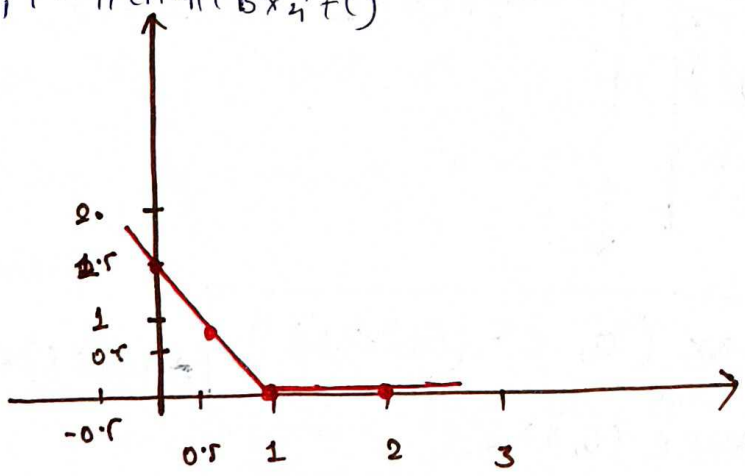
$m(0, 0.5)=0.5$

$Ax+by+c=-0.5$ pt (0.5, 0.5)

$\max(0, 1.5)=1.5$

pt (-0.5, 1.5)

$\max(0, 1-y_i(Ax_i+Bx_i+c))$



$y_i(Ax+by+c)$