Page No.
Date

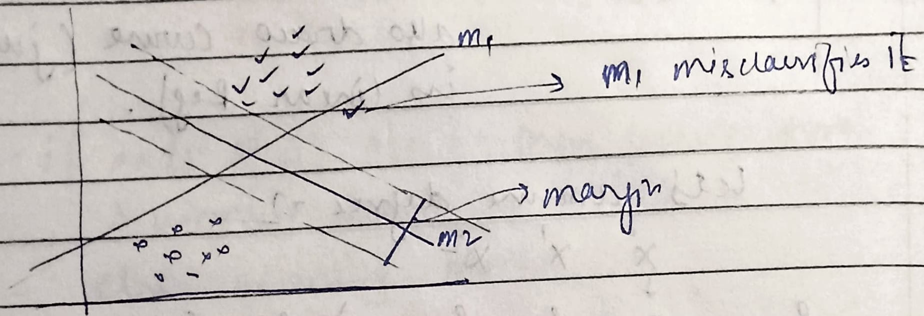
· SVM part-1
↳ It is generally used in-
→ classificat" (binary, multiclass)
→ Regression
→ Image processing

→ It can also be used on Non-linear Data set

· Maximal Margin Classifier - (Hard Margin SVM)

cgpa | iq | Place.

× → Not placed
✓ → Placed



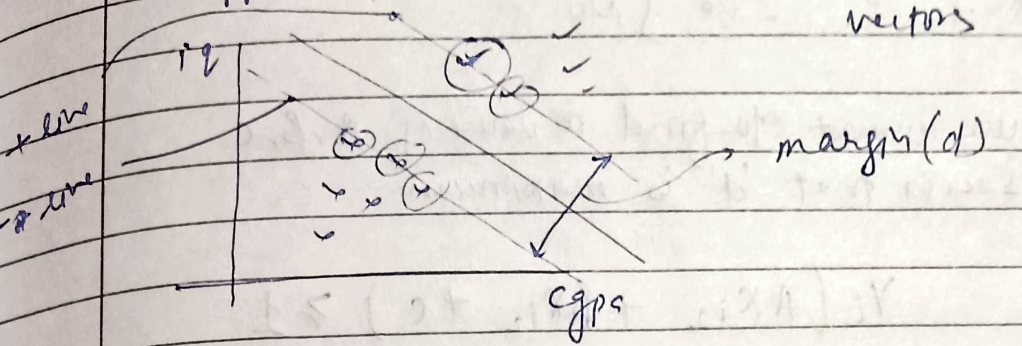
→ m_1, m_2 both can classify the classes but here m_2 is better model than m_1 because m_2 is creating a margin whereas m_1 is passing very close to pts. So its possible for new pts. that m_1 can misclassify

So SVM has more margin.

· The basic requirement is classes should be linearly separable

· Support vectors ·

These pts. are support vectors



→ The points which lie on the margin are support vectors

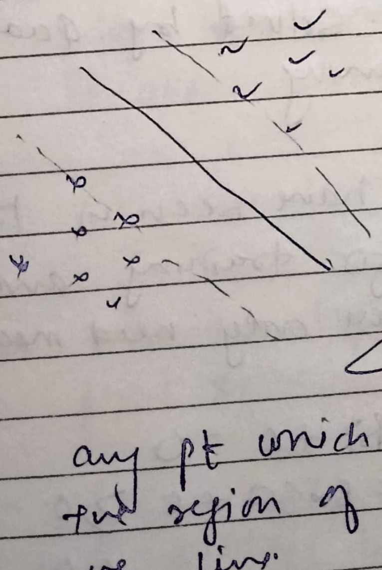
· <u>Mathematical</u> <u>formulation</u> –

↳ we want a line segment which maximize 'd' such that no points lie inside the margin.

Eqn of line : $Ax + By + C = 0$

" " +ve line : $Ax + By + C = 1$

-ve line : $Ax + By + C = -1$



Now we want to write Eqn for the condn

+ve line :- $Ax + By + C \geq 1$

-ve line :- $Ax + By + C \leq -1$

↳ By using it we can ensure any pt which come will either lie in +ve region of +ve line or -ve ~~+ve~~ region of -ve line.

Page No.
Date

→ $x = 1$ means +ve } Yes
$x = -1$ " -ve } NO

→ we want to find value of A, B, C
such that 'd' is maximum

$$Y_i (A X_{i_1} + B X_{i_2} + C) \geq 1$$

here $Y_i = 1$ means +ve or True
$Y_i = -1$ " -ve or False

↳ we know that the distance b/w two parallel line is $\frac{|c_1 - c_2|}{\sqrt{a^2 + b^2}}$

So $\frac{|c+1 - c+1|}{\sqrt{A^2+B^2}} = \frac{2}{\sqrt{A^2+B^2}}$

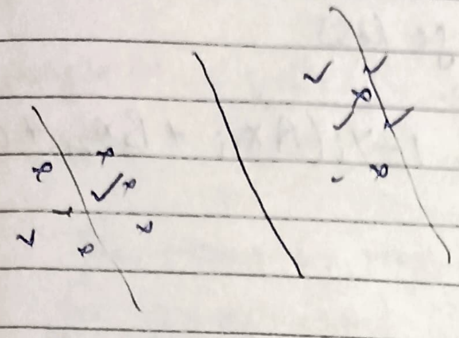
$\Rightarrow \underset{A,B,C}{\text{argmax}} \frac{2}{\sqrt{A^2+B^2}}$ given $\{ Y_i (A X_{i_1} + B X_{2_i} + C) \geq 1$

↳ It will be solved by Quadratic programming

→ All the above we have seen is training. we need margin for training and at time of prediction we only need model.

for prediction $Ax + By + C = 0$
$(8, 80)$ $= A \times 8 + B \times 80 + C \geq 0$ ✓
≤ 0 ✗ (No placement)

- Problem with Hard margin svm



→ can't apply here.

- Session-2

- in hard margin svm

$$\underset{A\,B\,C}{\arg\max} \; \frac{2}{\sqrt{A^2+B^2}} \quad \text{such that} \quad \underbrace{Y_i(AX_{1i} + BX_{2i} + C) \geq 1}$$

due to this constraint no red pt (x) can go in region of green point (✓) area. So we have to make this constraint flexible i.e soften it.

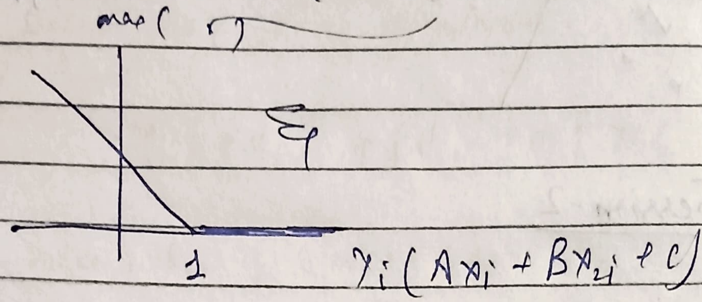
- Slack variable -
  - It is used to handle the cases where data are not linearly separable. It allows some degree of error in classificat".
  - for each data pt. i we ~~calculate~~ introduced $\xi_i \geq 0$

ξ_i measures degree of misclassificat" for data point x_i

- $\xi_i = 0$ x_i is on correct side of margin
- $0 < \xi_i < 1$ x_i is on correct side but on wrong side of margin
- $\xi_i \geq 1$ x_i is on wrong side

ξ_i → missclassification score
↳ also called hinge loss

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



→ Soft margin Svm –

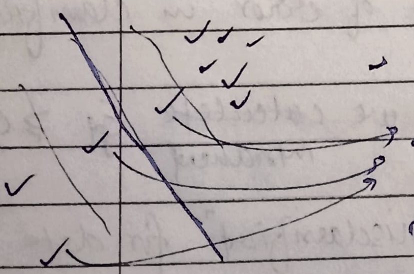
In hard margin svm $\rightarrow \underset{A,B,C}{\arg\max} \ \frac{2}{\sqrt{A^2+B^2}}$ such that ___

⇓

$\underset{A,B,C}{\arg\min} \ \frac{\sqrt{A^2+B^2}}{2}$ such that $y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 - \xi_i$

for all x_i such that $\xi_i \geq 0$

Here we have now introduced ξ_i



- After introducing the slack variable

These pts also starts following the constraints

i.e the constraint is followed by all the pts

means it is no more a constraint

This is not good so,

Page No.
Date

& we update it

Soft margin SVM

$$\underset{A,B,C}{\text{argmin}} \; \frac{\sqrt{A^2+B^2}}{2} + \frac{1}{n}\sum_{i=1}^{n} \xi_i \quad \text{Such that} \; -$$

This term is responsible for maximizing 'd'

it tells us total misclassification, So we want to minimize it

Due to this we will choose an optimum pt.

→ we can introduce a hyper parameter 'C'

$$\underset{ABC}{\text{argmin}} \; \frac{\sqrt{A^2+B^2}}{2} + C\,\frac{1}{n}\sum_{i=1}^{n} \xi_i \quad \text{such that}$$

if 'C' is more then our focus will be more on minimizing $\sum \xi_i$

& if C is less our focus will be more on minimizing $\frac{\sqrt{A^2+B^2}}{2}$

→ Bias variance tradeoff :-

if C value is high then overfitting (low bias high variance)

if C value is low then underfitting (high bias & low variance)