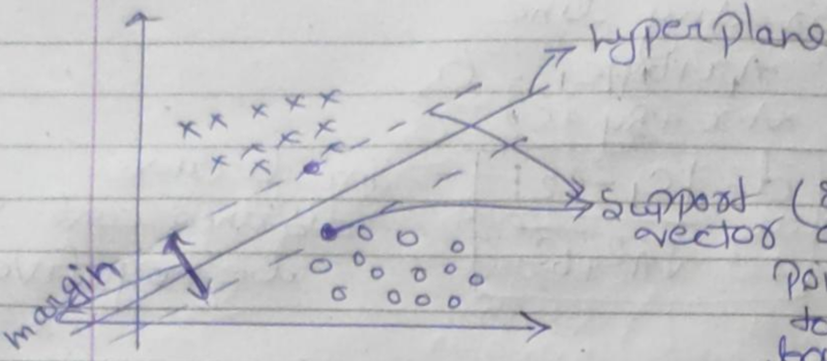
Support vectormachine.
↳ for classification



(support vectors are the data point closest to the decision boundary of classifier)

Hard margin
scenerio whor decision boundary strictly separates classes without allowing any misclassification.

Soft margin
here a margin of tolerance is allowed for misclassification.

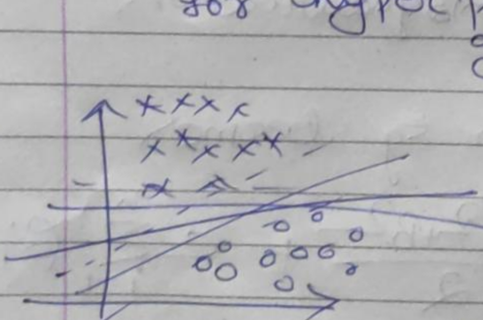
for two points:- Eq of lines

line $y = mx + c \Rightarrow ax + by + c = 0$

or $w_1x_1 + b$

for hyperplane

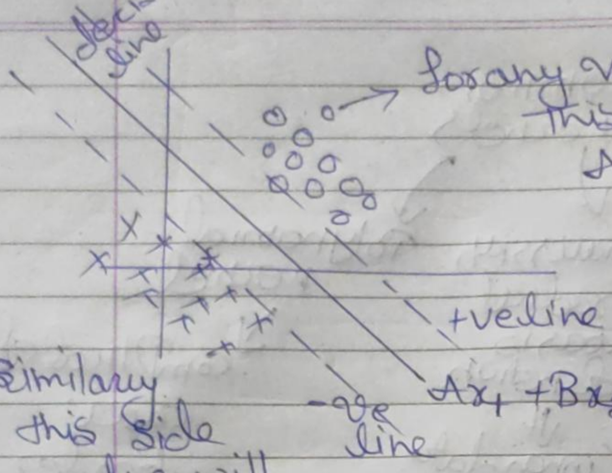
$$y = w^T x + b$$



In thes three which line is best fit because all three classifies well. In SVM we choose line for which distance of margin is maximum.

No.

Date : / /



for any value this side of line $Ax_{1i} + Bx_{2i} + C > 0$

Similarly this side value will be (-ve) $Ax_{1i} + Bx_{2i} + C < 0$

So इन +ve & -ve line को define कर सकते है।

by $Ax_{1i} + Bx_{2i} + C \geq K$ (we assume $K = 1$)

Then $Ax_{1i} + Bx_{2i} + C \geq +1$ —— (1)

Similarly for negative line

by $Ax_{1i} + Bx_{2i} + C \leq -1$ —— (2)

can we combine this both Eq.

~~Yes~~ because y_i also +ve for line of side +ve & -ve ~~at the~~ line side -ve.

$$\boxed{y_i(Ax_{1i} + Bx_{2i} + C) \geq 1}$$

condition

How to find distance:

↳ by line equation

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

(3)

No.

Date: / /

for Any line

$A_1x + By + C_1 = 0$

$Ax + By + C_2 = 0$

$$\boxed{d = \frac{|C_1 - C_2|}{\sqrt{A^2+B^2}}}$$ → mod because distance can't be negative

and SVM we try to maximize this distance d

$$\Rightarrow \boxed{\begin{array}{l} \arg\max \frac{2}{\sqrt{A^2+B^2}} \\ \text{for } y_i(Ax_{1i} + Bx_{2i} + C) \geq 1 \end{array}}$$

↳ That is the eq for hard margin SVM, here we can not misclassify any single point.

to overcome this problem we add a parameter ξ (misqualification score).

for soft margin eq becomes

$$\arg\min_{(A,B,C)} \frac{\sqrt{A^2+B^2}}{2} + C \sum_{i=1}^{n} \xi_i$$

Here C is an Hyperparameter. known as regularization parameter.

No.
Date: / /

⇒ It (C) controls the trade-off between maximizing the margine & minimizing the classification error.

⇒ A large value of C, narrower the margine & few misclassification, which may lead to overfitting.

⇒ A small value of C wider margine & may tolerate more misqualification prevents overfitting.

What is kernel trick?

↳ The "kernel Trisk" is a method is support vector machines (SVMs) which is used to convert data (that is not linarly separable) into a higher dimensional features space where it may be linearly separed.

This technique enables the SVM to identify a hyperplane that separates the data with the maximum margin. even when the data is not linearly separable in its orginal space

No.

Date: / /

most common kernels

linear kernel

(when data is approximately linear)

$K(x,y) = x^T y$

Gaussian (radial basis function)

(when data has no clear boundary & has many complicated areas)

$K(x,y) = \exp(-\gamma \|x-y\|^2)$

Polynomial kernel

complicated curve border.

$K(x,y) = (x^T y + c)^d$