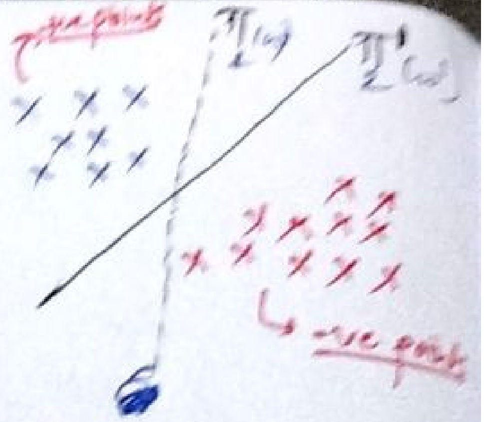
SVM [Support Vector Machines]

Geometric Intuition:

$\boxed{\pi: \text{Hyperplane}}$



⇒ Separate +ve from -ve point as widely as possible

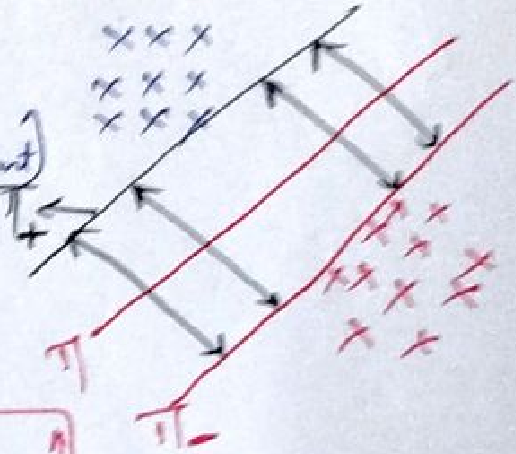
π: Margin Maximizing hyperplane

SVM: find a π that maximized the Margin = dist(π_+, π_-)

$\pi_+ \parallel \pi \parallel \pi_-$

[HP $\parallel$ to π that touches first +ve-point go towards +ve points]

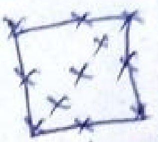
Sim π_-



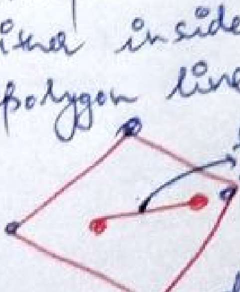
⇒ $\boxed{\text{Margin} \uparrow \implies \text{Generalize accuracy} \uparrow}$

Alternative Geo.

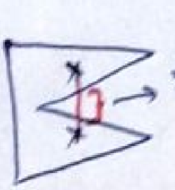
↳ Convex-hull:



Convex-polygon
सारे points को cover
either inside or in
polygon line itself.

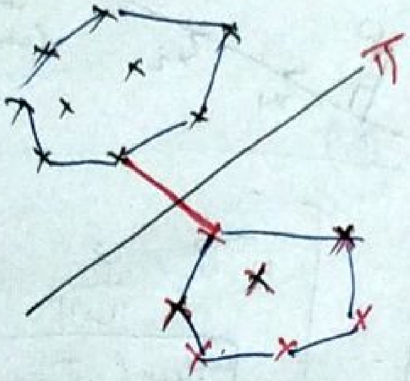


किसी भी
दो points को
meet पर smallest
dis is ~~...~~ inside the polygon



ये part बाहर जा रहा है
so ~~not convex~~
non-convex polygon

Alternative:



① Find convex hull for both ~~[illegible]~~ (+ve) & (−ve) points

② Join shortest line connecting the hulls

③ Bisect the line (π)

मिलाणा

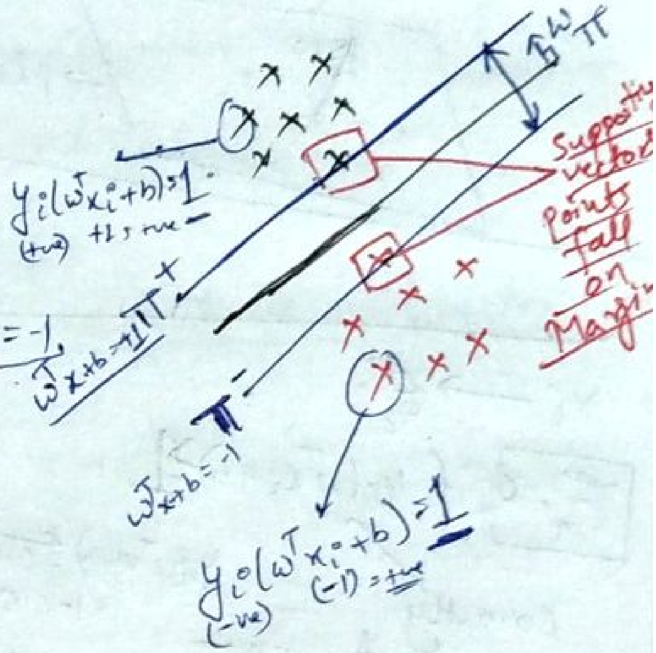
Mathematical →

$w \perp$ to π since $\pi_+ \parallel \pi \parallel \pi_-$

so $w \perp$ to π_+ & π_-

$\pi = w^T x + b = 0$

$\pi_+ : w^T x + b = +1$ | $\pi_- : w^T x + b = -1$



Margin $d = \frac{2}{\|w\|}$ → norm

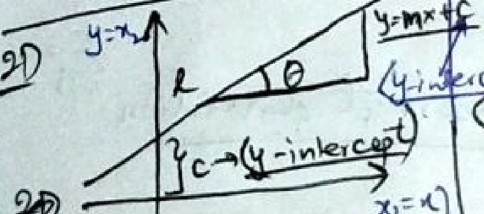
→ (w^*, b^*) argmax $\frac{2}{\|w\|}$

Line: $y = mx + c$

$ax + by + c = 0$

$y = -\frac{c}{b} - \frac{a}{b} x$

2D



$w_1 x_1 + w_2 x_2 + w_0 = 0$

$x_2 = -\frac{w_0}{w_2} - \frac{w_1 x_1}{w_2}$

c

n-D hyperplane

$w_0 + \sum_{i=1}^{n} w_i x_i = 0$

$w_0 + [w_1 \; w_2 \; \ldots \; w_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 0$

$w_0 + w_1 x_1 + w_2 x_2 + \ldots w_n x_n = 0$

Vector: $\boxed{w_0 + w^T x = 0}$

$\pi_n : w^T x = 0$ $w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

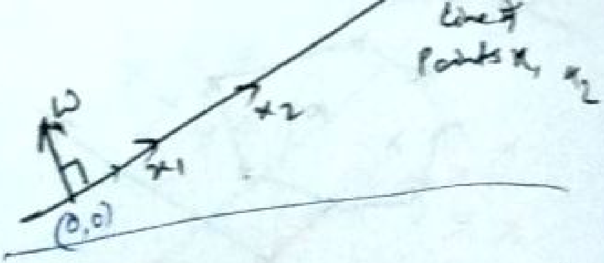
$\boxed{w \cdot x = w^T x = \|w\| \|x\| \cos\theta} = 0$

$(\theta = 90)$ i.e. $\boxed{w \perp x}$

→ $w^T x = 0$

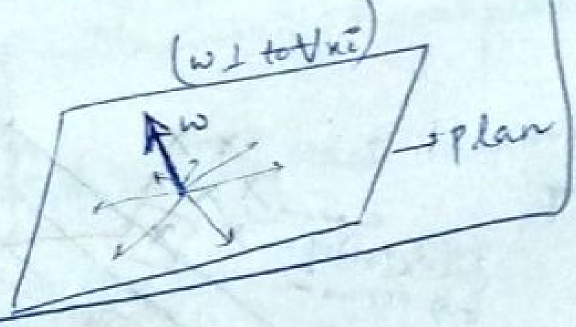
→ w is $\perp$ to π-plane

$\hat{w} = \frac{w}{\|w\|}$



$\boxed{\hat{w} = \frac{w}{\|w\|}}$

$\boxed{\hat{w} \cdot x_i = 0 \quad \forall x_i \in \pi}$



$w^T, b^* = \frac{2}{\|w\|}$

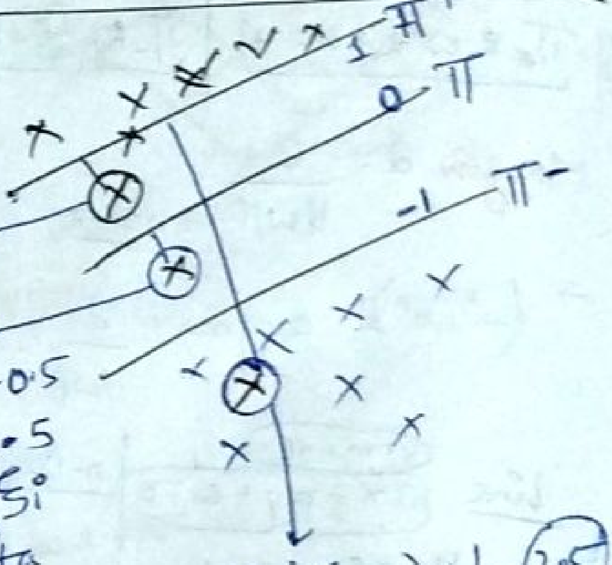
$\boxed{\forall y_i (w^T x_i + b) \geq 1}$

data is linearly separable to π_+ & π_-

$x_i \longrightarrow \xi_i$

$\boxed{\xi_i = 0}$, if $y_i(w^T x_i + b) \geq 1$

Correctly Classified π^+ & π^-



$y_i(w^T x_i + b) = 0.5$
$= 1 - 0.5$

$y_i(w^T x_i + b) = -0.5$
$1 - (1.5)$
ξ_i
zeta

$y_i(w^T x_i + b) = 1 - 2.5$
ξ_i

~~ξ_i~~

$\xi_i > 0$: ~~if $(y_i(w^T x_i + b))$~~
↳ some unit of distance away from correct hyperplane in incorrect direction

Think Main काम था Maximize $\frac{2}{\|w\|}$ Margin

सोच Max. $f(x)$ mea → Minimize $\frac{1}{f(x)}$

(w^*, b^*) Max $\rightarrow \frac{2}{\|w\|} \longrightarrow$ ny Min $\frac{\|w\|}{2}$

$$\boxed{(w^*, b^*) = \underset{w, b}{\operatorname{argmin}} \frac{\|w\|}{2} + C \cdot \frac{1}{n} \sum_{i=1}^{n} \xi_i}$$

→ avg dis. of mis classified points from π

Aim: $\boxed{\text{Minimize Errors}} \rightarrow \boxed{\text{Min. Misclassifications}}$

$$(w^*, b^*) = \operatorname{argmin} \frac{\|w\|}{2} + C\left(\frac{1}{n}\sum \xi_i\right)$$

(regularizator)

hingeloss

loss to model

hyperparameter

Soft-Margin

C ↑ : Tending to make mistakes ↓ on D_{Train} ☆

C ↓ : underfit → highly bias

~~Logistic Regression~~

$\|w\| \neq 1$

↳ Need not be unit vector

$$\boxed{y_i(w^T x_i + b) \geq 1 - \xi_i} \quad \forall \ \xi_i \geq 0.$$

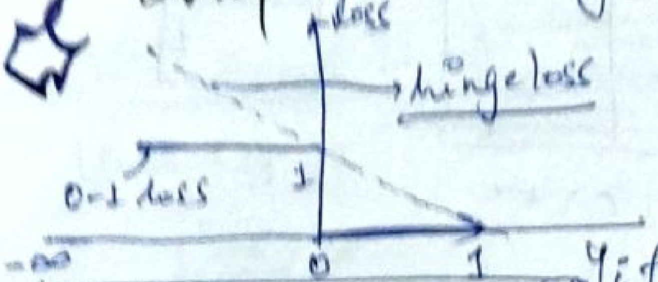
~~if cnt ≤ ||~~

SVM

① Logis. Reg. ⇒ Logistic Loss + Regularization

Linear Reg ⇒ Linear loss + Reg.

SVM ⇒ Hinge Loss + Reg.



$y_i f(x_i)$

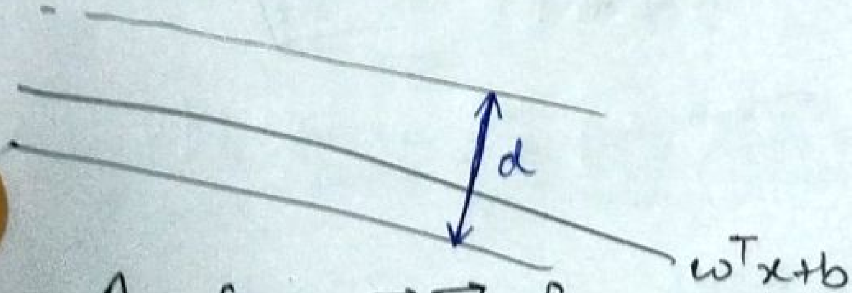
$\boxed{\begin{cases} z_i > 0 : x_i \text{ correctly classif} \\ z_i < 0 : x_i \text{ Incorr clas} \end{cases}}$ $y_i(w^T x_i + b) = z_i$

Hinge loss: $\begin{bmatrix} z_i \geq 1 ; \text{hinge-loss} = 0 \\ z_i < 1 ; \text{hinge loss} = 1 - z_i \end{bmatrix} \rightarrow$ ma

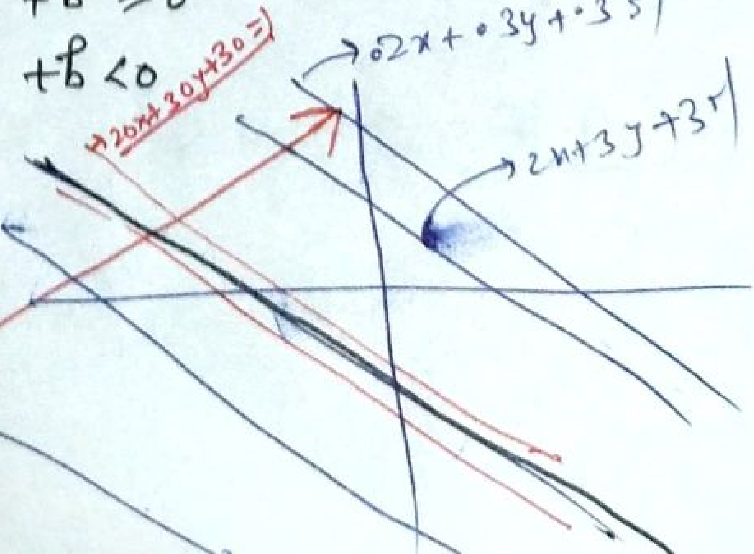
$\rightarrow \max[0, 1 - z_i]$

Case 1: $z_i \geq 1 \Rightarrow \boxed{1 - z_i : -ve} \longrightarrow 0$

$z_i < 1 \quad 1 - z_i > 0 \longrightarrow \max(0, 1 - z_i)$



$$\hat{y} = \begin{cases} +1 & \vec{w} \cdot \vec{x_i} + b \geq 0 \\ -1 & \vec{w} \cdot \vec{x_i} + b < 0 \end{cases}$$



$2x + 3y + 3 = 1 \quad -1$

$20x + 30y + 30 = 1 \quad -1$

$.2x + .3y + .3 = 1 \quad -1$

Margin 10 के Factor मे है तो compress → Contract/shrink
Margin $\frac{1}{10}$ के Factor मे to expand

d_i के ऊपर constraint है

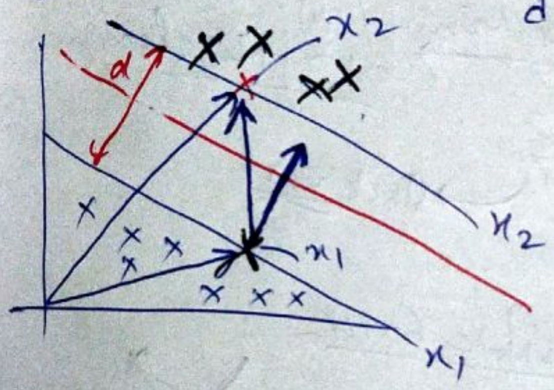
+ve points (−ve) point की तरफ नहीं आने चाहिए

$\boxed{y_i(\vec{w}\cdot\vec{x_i}+b) \geq 1}$ → equal for support vector on line

[determine point is above Π_+ or Π_-]

$$\boxed{y_i[\vec{w}\cdot\vec{x_i}+b] \geq 1}$$

Aim: Find d



$$d = (\vec{x_2}-x_1)\cdot\frac{\vec{w}}{||w||} = \frac{(\vec{x_2}\cdot\vec{w})-(\vec{x_1}\cdot\vec{w})}{||w||}$$

$$\boxed{y_i(\vec{w}\cdot\vec{x_i}+b)=1}$$

$$1(\vec{w}\cdot\vec{x_2}+b)=1 \quad \Big| \quad \vec{w}\cdot\vec{x_2} = 1-b$$

Sim: $\cancel{-1}(\vec{w}\cdot\vec{x_1}+b) = \cancel{-}1$

$$\vec{w}\cdot\vec{x_1}+b = -1$$

$$\vec{w}\cdot\vec{x_1} = -b-1$$

$$d = \frac{(\vec{x_2}\cdot\vec{w})-(\vec{x_1}\cdot\vec{w})}{||w||} = \frac{(1-b)-(-b-1)}{||w||} = \frac{2}{||w||}$$

Harg $\underset{(w^*,\,b^*)}{\arg\max}\ \frac{2}{||w||}$ s.t. $\boxed{y_i(\vec{w}\cdot\vec{x_i}+b)\geq 1}$

Hard Margin SVM

Hard Margin Margin Maximum करना है then we have to

Maximize $\frac{2}{||w||}$ st. constr $y_i(w^Tx_i+b) \geq 0$

-ve point π_- के ऊपर नहीं जाएगा towards

और vice versa

getting such case is not real

अब हमें इसे ऐसा modify करना पड़ेगा कि ये outlier को रहने की जगह दे

Soft Margin Maximize f(x) → Minimize f(x) → Cost of misclassification

$$\text{argmin} \; \frac{||w||}{2} + C\sum_{i=1}^{n} \xi_i$$

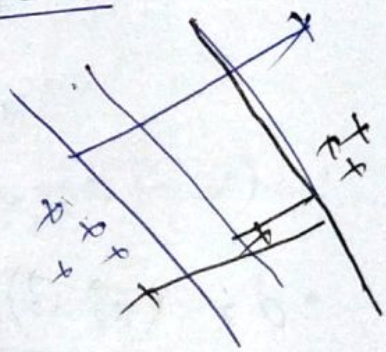
ξ_i → slack var

C ↑ hyperparam

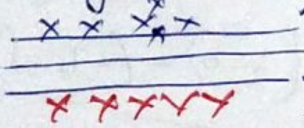
SVM Error

Margin Error + Classific Error

Classific Error → सोच ξ

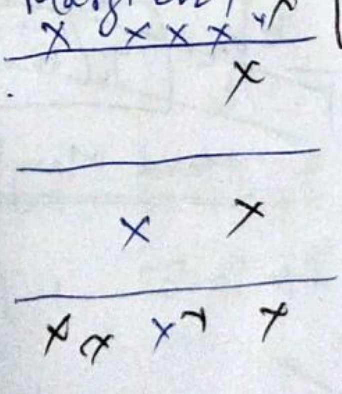


Class↑ 'Margin error कम

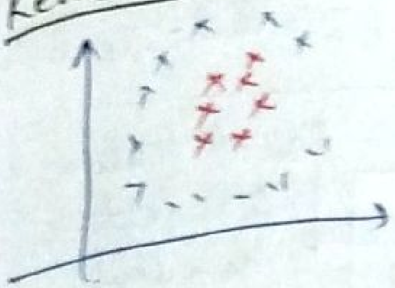


गलती कम overfit

ClassErr↓ Margin err↑ [गलतीयाँ ज्यादा



Kernel Trick:



RBF
Polynomial
Sigmoid

RBF: Radi Basx^n e^{-x^2}

Each vector is augment with 1 as bias

$$\begin{bmatrix} \alpha_1 \hat{s_1}\cdot\hat{s_1} + \alpha_2 \hat{s_2}\cdot s_2 = -1 \\ \alpha_1 \hat{s_1}\cdot s_2 + \alpha_1 s_2 \cdot s_2 = 1 \end{bmatrix} \rightarrow \begin{bmatrix} 3\alpha_1 + 5\alpha_2 = -1 \\ 5\alpha_1 + 9\alpha_2 = 1 \end{bmatrix} \rightarrow \boxed{\begin{matrix} \alpha_1 = -7 \\ \alpha_2 = 4 \end{matrix}}$$ ✓ ✓

SVM: $\bar{w} = \sum_{i=1} \hat{\alpha_i}\hat{s_i} = 7\begin{bmatrix}1\\1\\1\end{bmatrix} + 4\begin{bmatrix}2\\2\\1\end{bmatrix} = \begin{bmatrix}1\\1\\-3\end{bmatrix} \rightarrow w\begin{pmatrix}1\\1\end{pmatrix}$

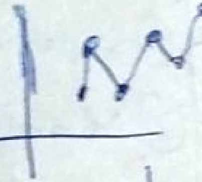
$\alpha_1 s_1 \quad \alpha_2 s_2$

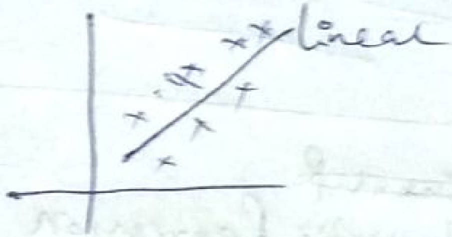
$\underline{b = -3}$

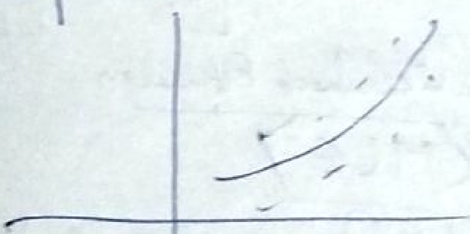
Regularization

$\rightarrow z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \ldots$

(overfitting)







Eliminate some features

Nullify effect on certain parameters

Cost $\frac{1}{m}\sum_{i=1}^{m} Loss + \sum_{j=1}^{n} \frac{\lambda}{2m}|w_j|^2 \quad w\downarrow \quad w_i \approx 0$