SVM AND SVR Algorithms

① CLASSIFICATION AND REGRESSION PROBLEMS.

ML Breakup

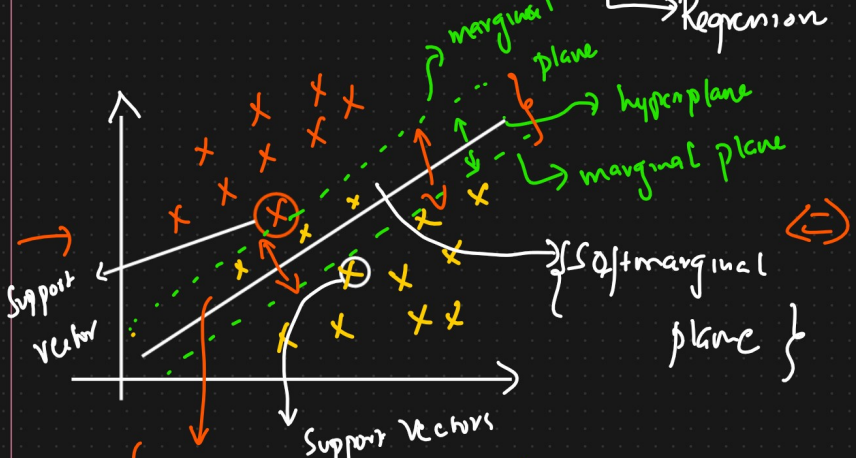
↓

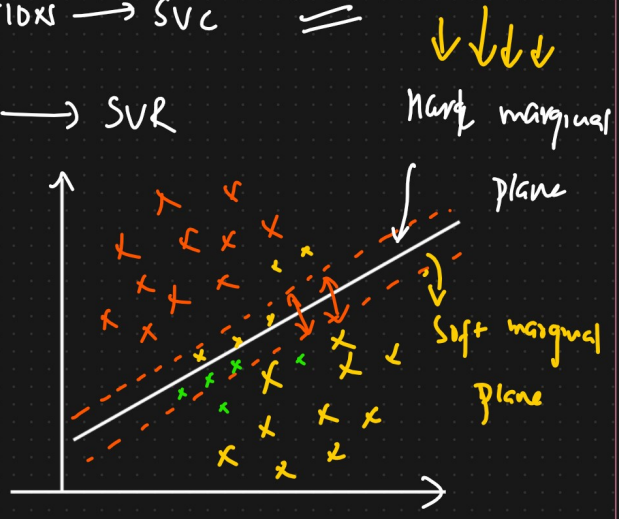
SVM

{AIM}

Support Vector Machine → CLASSIFICATION → SVC

→ Regression → SVR





{Hard marginal plane}

$y = mx + c$

$y = \theta_0 + \theta_1 x$

$y = \beta_0 + \beta_1 x$

$y = w^T x + b$

$ax + by + c = 0$ → Equation of a straight line

$by = -ax - c$ → coefficient

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

$y = w_1x_1 + w_2x_2 + w_3x_3 + b$ {Matrix multiplication}

$y = w^T x + b$

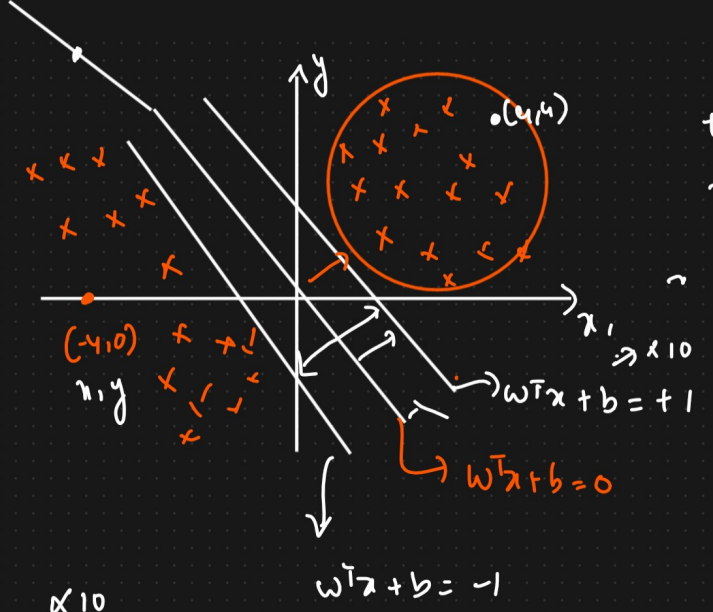
$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$

⇕

$y = m_1x_1 + m_2x_2 + m_3x_3 + c$

$y = m^T x + c$

$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$



k=1

+1

-1

$3x + 2y + 4 = 0$

$3(4) + 2(4) + 4 = 0$

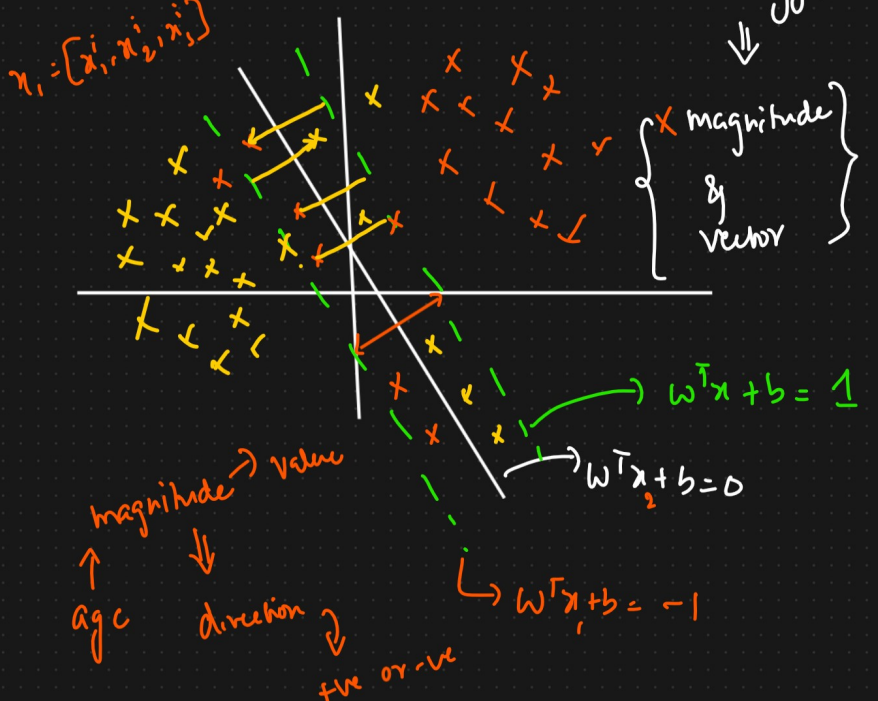
$12 + 8 + 4$

$\Rightarrow 24 \Rightarrow$ +ve

$3x + 2y + 4 = 0$

$3(-4) + 2(0) + 4 = 0$

$-8 \Rightarrow$ -ve value



w ⇒ coefficient

⇓

magnitude ⇒ value

⇓

age direction

⇓

+ve or -ve

$w^T x_1 + b = +1$

$w^T x_2 + b = -1$

(-) (-) (+)

$\vec{w}$

$\frac{w^T (x_1 - x_2)}{\|w\|} = \frac{2}{\|w\|}$ ↑↑↑.

Marginal Error

Maximize (w, b) $\frac{2}{\|w\|} \Rightarrow$ Marginal plane ✓

Constraint

Such that

Truth o/p

$y_i = \begin{cases} +1 & w^T x + b \geq 1 \rightarrow \text{correct} \\ -1 & w^T x + b \leq -1 \rightarrow \text{+ve} \end{cases}$

For all correct points

Constraint → $y_i * (w^T x_i + b) \geq 1$

Maximize $\frac{2}{\|w\|}$ (w,b) $\Leftrightarrow$ Min $\frac{\|w\|}{2}$ (w,b)

Linear Regression

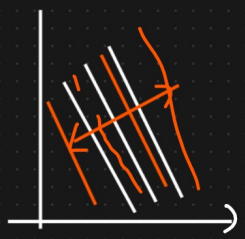
Cost function : Min $\frac{\|w\|}{2}$ (w,b) $+ C_i \sum_{i=1}^{n} \xi_i$

C_i → Hyperparameter

ξ_i → GTA

→ {Hinge Loss}

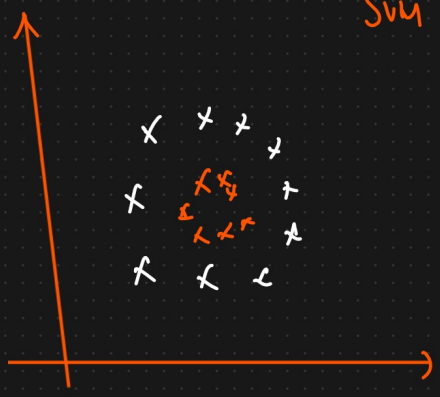
→ Summation of the distance of the wrong data points from the marginal plane



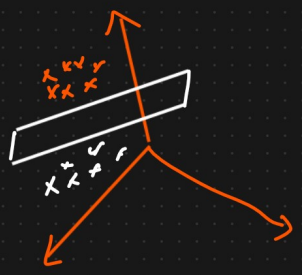
{ Hard marginal plane

{ How many points we want to avoid misclassification.

Svm



Svm Kernels (Kernel Trick)

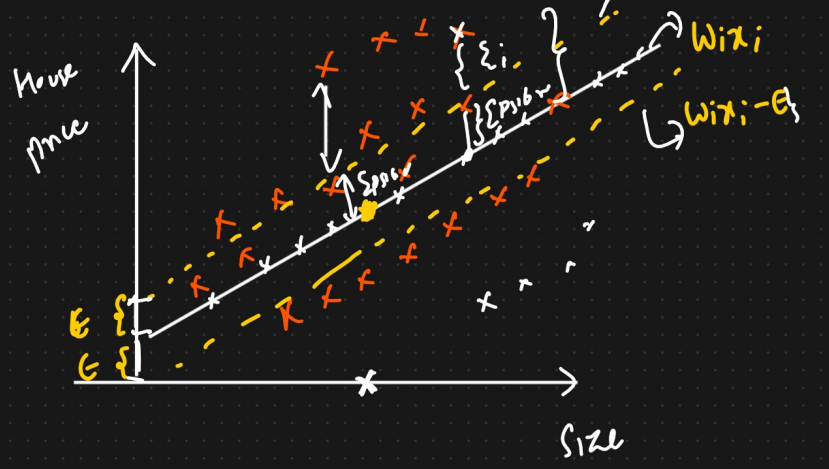


⑥ Support Vector Regression

$w_i x_i + b = +1$

$w_i x_i + b = -1$

$\{ w_i x_i + \epsilon \}$



$$\underset{(w,b)}{\text{Min}} \quad \frac{\|w\|}{2}$$

Constraint

$$\left.\begin{array}{l} |y_i - w_i x_i| \leq \epsilon \end{array}\right\}$$

$\hat{y}$

Cost function

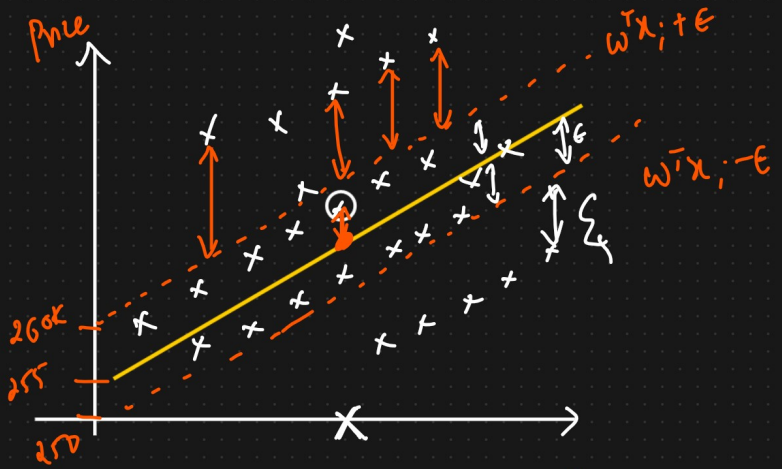
$$\underset{(w,b)}{\text{Min}} \quad \frac{\|w\|}{2} + C \sum_{i=1}^{n} |\xi_i|$$

Constraint

$$|y_i - w_i x_i| \leq \epsilon + |\xi_i|$$

ϵ : Epsilon

$|\xi_i|$: distance of the points that are away from the Epsilon line



Hyperparameter

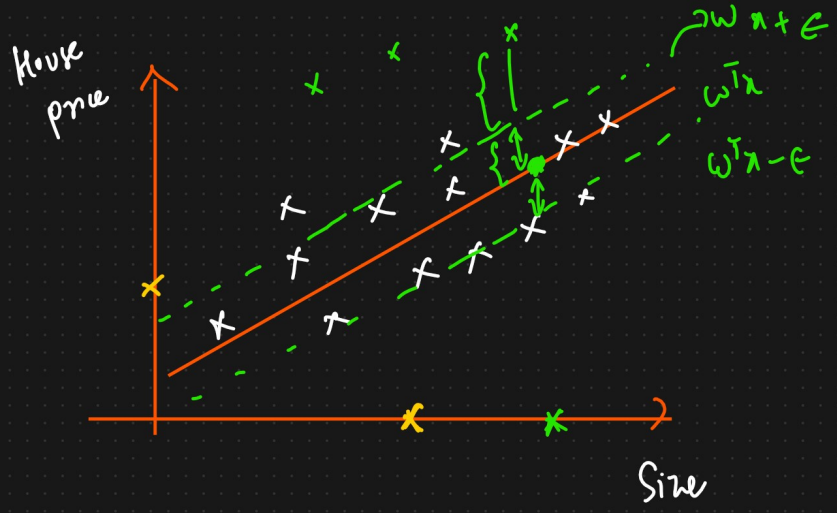
$$\underset{(w,b)}{\text{Min}} \quad \frac{\|w\|}{2} + C \sum_{i=1}^{n} \xi_i$$

Constraint

$$|y_i - w^T x_i| \leq \epsilon + \xi_i$$

y_i : Truth point

$w^T x_i$: Predicted point



$|y_i - w^T x_i| \leq \epsilon + |\xi_i|$

$\hat{y}$

{ C & Slope }

Cost function = $\frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{i=1}^{n} (\text{slope})^2$

y_i & $w^T x + b$

$\hat{y}_i$

Actual value

predicted value

$\lambda \sum_{i=1}^{n} |\text{slope}|$

L2

L1

Regularization parameter

{ L2, L1, Elastic Net, Hinge Loss }

Answer: Relationship between C & λ

$= \quad =$

$C \propto \frac{1}{\lambda}$

Logistic Sklearn

Diff

Loss & cost function

1-datapoint

Loss fn = $\frac{1}{2}(y - \hat{y})^2$

$(\hat{y} - y)^2$

$(y - \hat{y})^2$

n-time datapoint

Cost funct: $\frac{1}{2n} \sum_{i=1}^{n} (y - \hat{y})^2$

fit & fit_transform & transform.

Standard Scaler

Transforming the data ??

Changing the data.

	f1	f2
Train	2	
	3	
	4	
Test	5	
	6	

f1 → f1-Standard

f1	f1-Standard
2	-1
3	0 ← μ, σ
4	1

Z-score = $\frac{x_i - \mu}{\sigma} = \frac{3}{1} = 1$

fit & transform.

fit & fit_transform

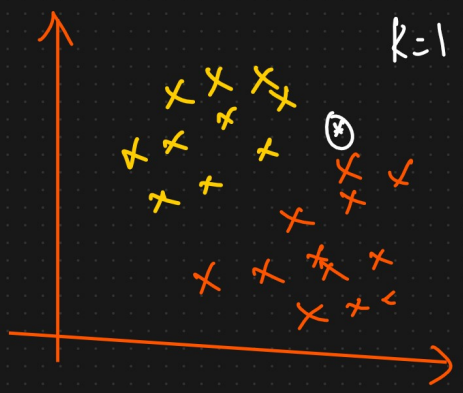
$\frac{2+3+4}{3} = \frac{9}{3} = 3$

Transform

Test

5
6

$\frac{5-3}{1} = 2$ No



K=1 bad accuracy.

K=2

K=3

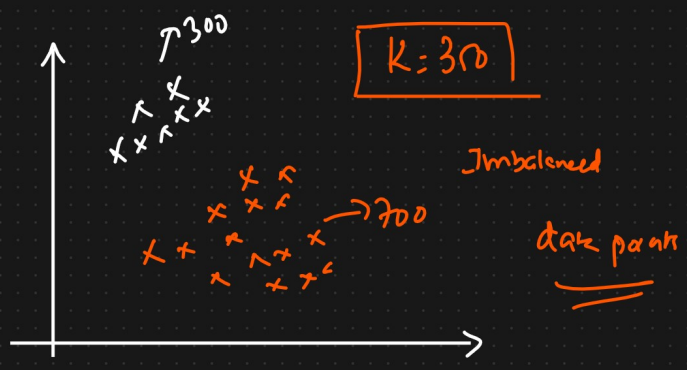
K=4

K=5

K=6

K=odd

K↑ Accuracy↑ Error↓



K = 30

Imbalanced

data point