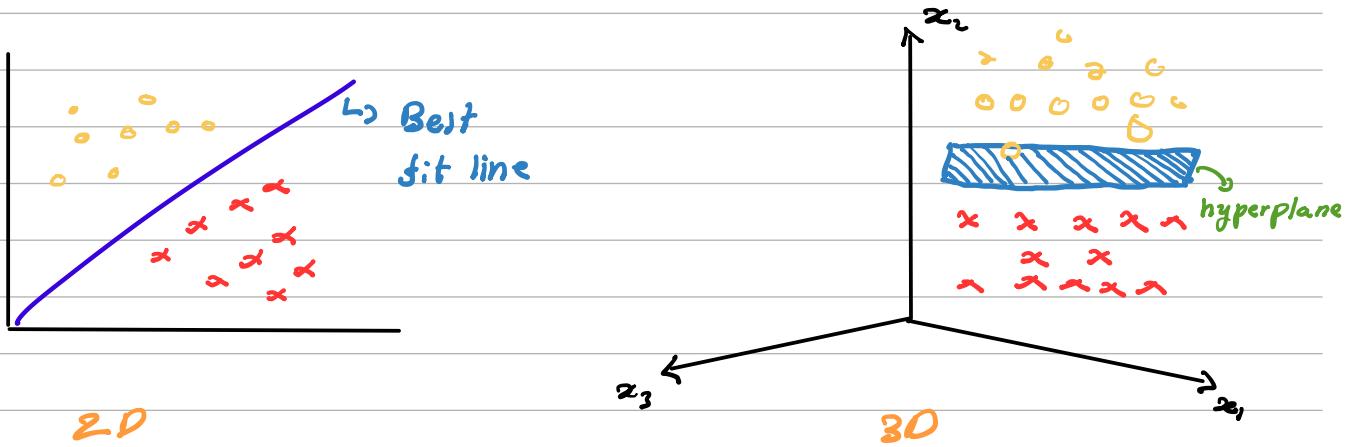


06 - Support Vector Machines

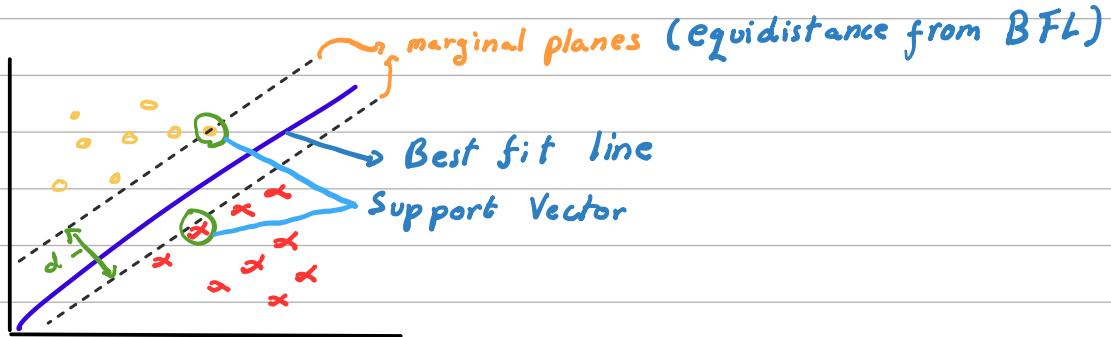
Advantage of SVM

- 1) Robust to outliers (able to handle outliers very well)
- 2) Non Linear data - by using kernels
- 3) Can use both classification & Regression Problem

- ① SVC (Support Vector Classifier) → Classification
- ② SVR (Support Vector Regressor) → Regression

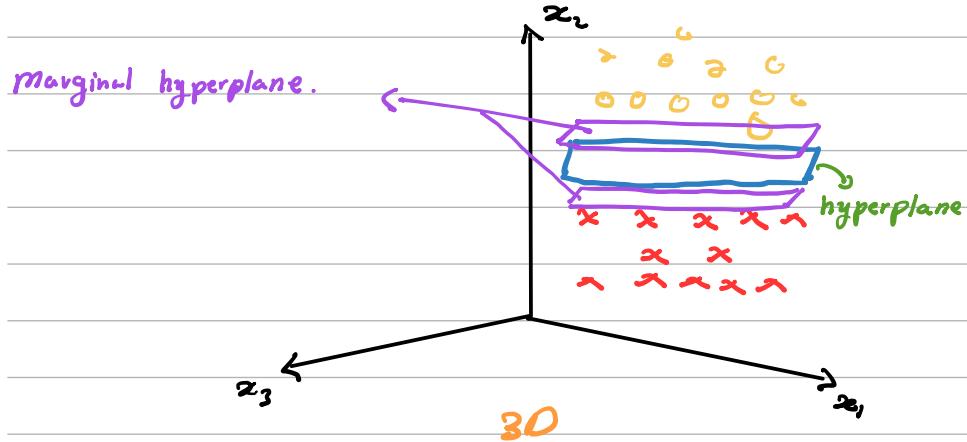


(1) Support Vector Classifier (SVC)



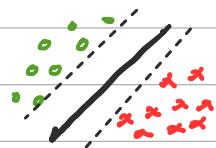
We draw 2 marginal line along with best fit line with these line
(d) distance should be maximum

The Marginal plane are equal distance from Best fit line.

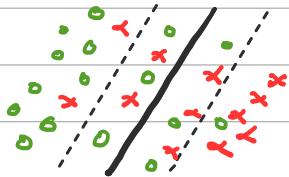


Soft margin & hard margin

hard margin: No points are missclassified, the points can easily be classified, no overlapping occurs



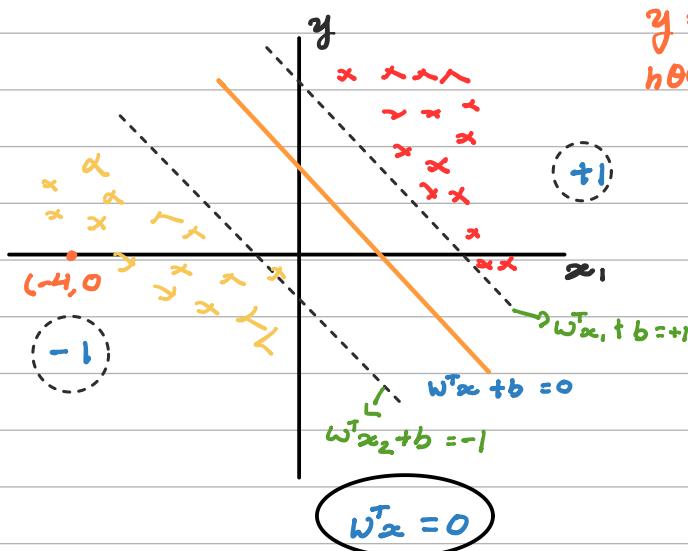
Soft margin: Some data points are missclassified, due to overlapping of data its hard to classify easily.



model knows has errors but predicts maximum data points.

In Real word dataset maximum we get soft margin dataset.

SVM math Intution



$$y = mx + c \iff ax + by + c = 0$$

$$h\theta(x) = \theta_0 + \theta_1 x_1$$

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$y = b + \omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3$$

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

$$w^T x = 0$$

if line passes
through Origin

$$\begin{aligned} w^T x_1 + b &= +1 \\ w^T x_2 + b &= -1 \\ (-) &\quad (-) \quad (+) \end{aligned}$$

$$w^T(x_1 - x_2) = 2$$

Cost function

$$\underset{w, b}{\text{maximize}} \frac{2}{\|w\|} \Rightarrow \text{Distance b/w MP}$$

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

\overrightarrow{w}

constraint such that

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

+1
-1

marginal
Plane

For all correct classified data points

$$y_i * [w^T x + b] \geq 1$$

Modified Cost function of SVC

$$\underset{w, b}{\text{maximize}} \frac{2}{\|w\|} = \underset{w, b}{\text{minimize}} \frac{\|w\|}{2}$$

Cost function of SVM (Soft margin)

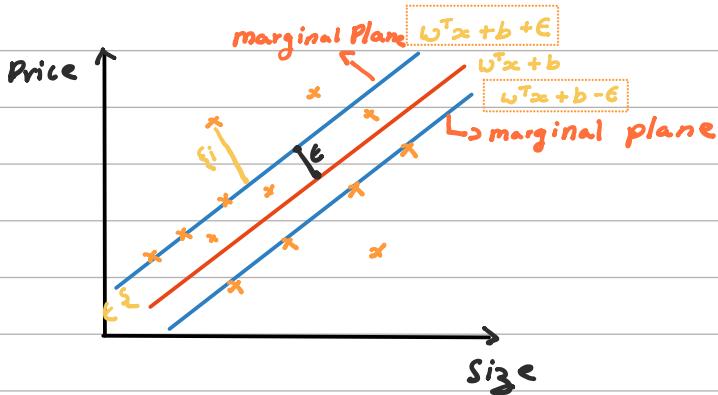
$$\text{Cost fun} = \min_{w,b} \frac{\|w\|}{2} + C \sum_{i=1}^n \xi_i$$

= hinge loss

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

{ How many points we can consider for misclassification?
 Hyperparameter }
 { Summation of the distance of incorrect data points from marginal plane }

② Support Vector Regressor (SVR)



$$\text{Cost fun} = \min_{w,b} \frac{\|w\|}{2} + C \sum_{i=1}^n \xi_i$$

= hinge loss

\Downarrow

Constraints

$$\text{Error} \Leftarrow |y_i - w^T x_i| \leq \epsilon + \xi_i$$

$\epsilon \Rightarrow$ marginal Error

$\xi \Rightarrow$ Error above the margin

SVM Kernels

