Date:

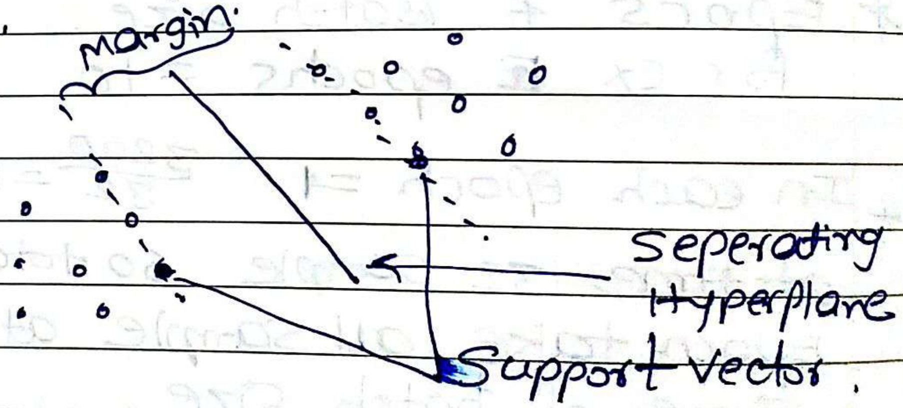
Support Vector Machine

* SVM can be used for classifier or regression challeges

* SVM tries to define hyperplane which can split the data in the most optimal way such that there is a wide margin among the hyperplane and observation.

* What is hyperplane
- Hyperplane is a generalization of a plane.
1) In one dimension, an hyperplane is point.
2) In two dimension it is line.
3) In three dimension it is plane.
4) In more dimension it is an hyperplane.

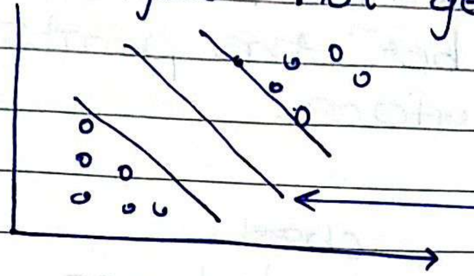
- Hyperplane categories data into different category.



* Purpose of SVM is to find optimal hyperplane so it needs to check how effective these hyperplanes.

* We need to choose correct hyperplane which will seperate data in optimal way.

* If hyperplane select which is close to the data points of one class. Then it meight not generalize well.



* To choose optimal hyperplane compute distance between hyperplane & closest data point.

* Advantage
1) Support vector machine have claver way to prevent overfitting.
2) SVM work with large no. of future without require too much computation.

Revision : Logistic Regression.

1) We have to find out probability $Y=1$.
   $P(Y=1|x)$ for $h(x)$
   Predict 1 when $h(x) > 0.5$
   Predict 0 when $h(x) < 0.5$
   Probability is higher when $h(x)$ is higher than 0.5.

2) Higher probability of conifidence output is higher when $h(x)$ is much larger than 0.5 because of sigmoid function.

→ As $h(x)$ is closer to one probabity or confidence is higher.
→ So more the confident which have further from the decision surface.

* SVM :-

1) SVM is classifier that maximizes the separation betn the points and the decision surface.

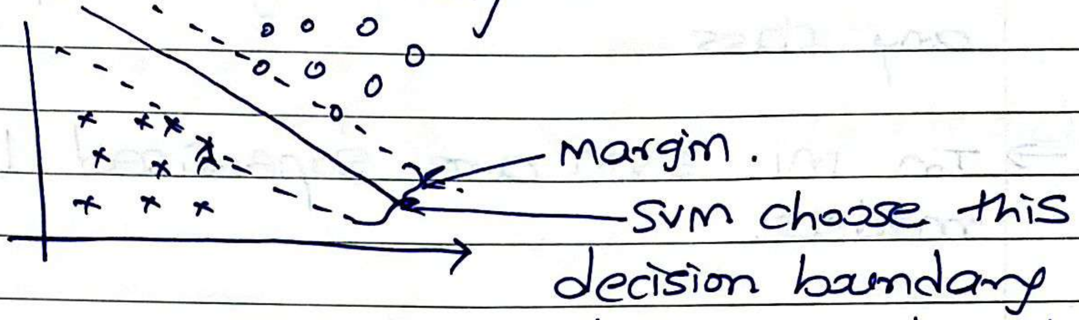
2) margin :-
closest
i) find the distance of this points from the decision surface.
ii) All possible decision surface we want to choose that one for which the margin width is highest.

3) Maximizing the margin width, the distance of the closest negative point to this line & the closest positive point to this line will be same & there are minimum of two support vectors or more but typically the number of support vectors is extremly small.

4) This support vectors are the one which actually determine by the equation of the line and typically the support vector number will be very small and this particular decision surface decides which point given a test point of decision surface based on classify the point has positive or negative.

* Compared to logistic regression and neural network, the support vector machine or SVM sometimes gives a cleaner and sometimes more powerful way of learning complex non-linear functions.

* SVM Decision Boundary



mathematically this selected decision boundary has a larger distance. This distance is called the margin.

The other two lines come close to the training example and seems not good to seperating the positive and negative classes than selected line.

* So this distance is called margin of ~~SVM~~ SVM & this gives the SVM a certain robustness, because it tries to separate the data with as a large margin as possible.

* So SVM is some times called a large margin classifier & need optimization problem.

Date:

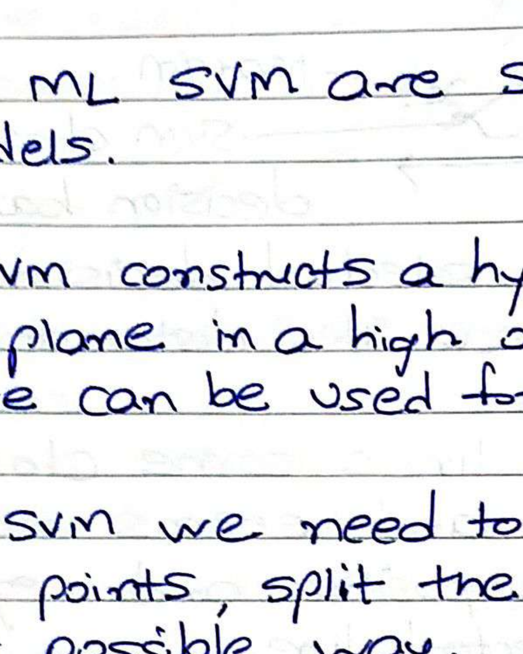
* SVM used in classification
Classification can be viewed as the task of seperating classes in feature space

→ Intuitively a good seperation is achived by the hyperplane that has the largest distance to the nearest training-data point of any class.

→ In ML SVM are supervised learning models.

→ A SVM constructs a hyperplane or set of hyperplane in a high or infinite dimensional space can be used for classification.

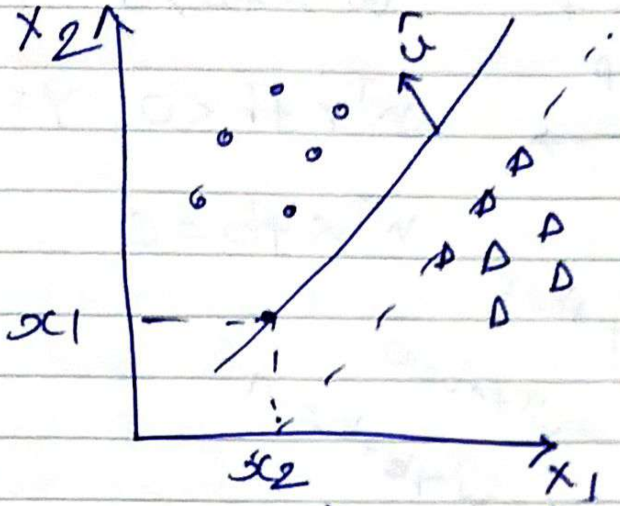
→ In SVM we need to find line which seperate data points, split the point in the best possible way.



* Margin is just the distance from the plane to a point.

Biggger margin able to achieve better out of sample performance.

* So in svm need to draw seperate line to seperate points.
Not only seperate points but also maximize margin bet^n them.



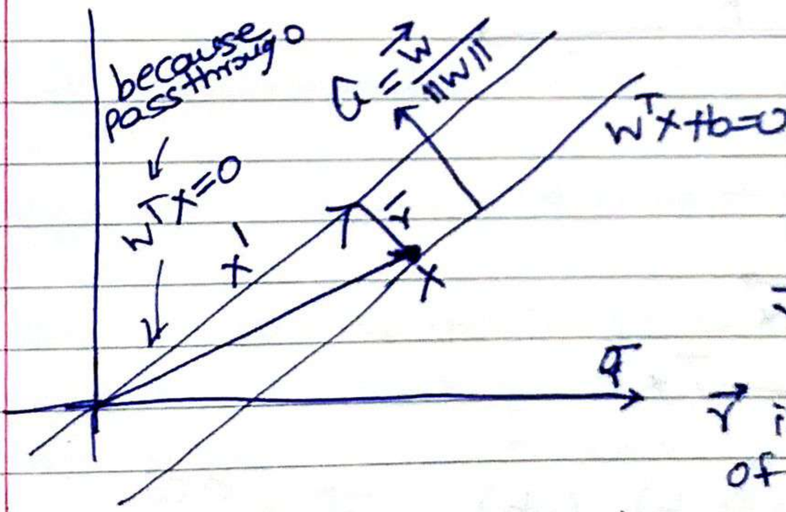
$$w^T x + b = 0$$
for $x \in R^2$

$$w_1 x_1 + w_2 x_2 + b = 0$$

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

$$\hat{u} = \frac{\vec{w}}{\|w\|_2} = \frac{\vec{w}}{\sqrt{w_1^2 + w_2^2}}$$

$\hat{u}$:- Unit vector on line.



$$\vec{r} = \vec{x'} - \vec{x}$$

$$\vec{r} = r\hat{u} = r \frac{\vec{w}}{\|\vec{w}\|_2}$$

$\vec{r}$ is along the direction of $\hat{u}$.

$w^T x = -b \qquad w^T x' = 0$

$\therefore w^T x' - w^T x = b$

$w^T (x' - x) = b$

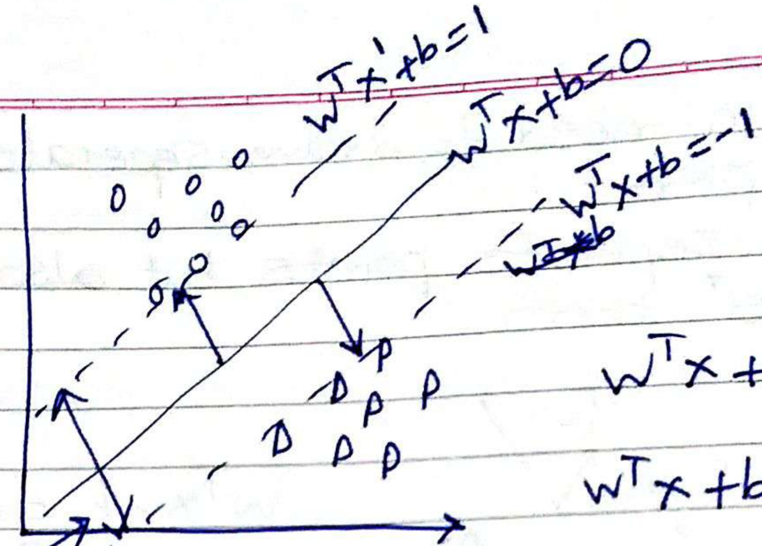
$w^T \vec{r} = b$

$r \frac{w^T w}{\|w\|_2} = b$

$$w^T w = w_1^2 + w_2^2 = \left(\sqrt{w_1^2 + w_2^2}\right)^2 = \|w\|_2^2$$

$\therefore b = \frac{r}{\|w\|}$

b denote how much need to move from origin.



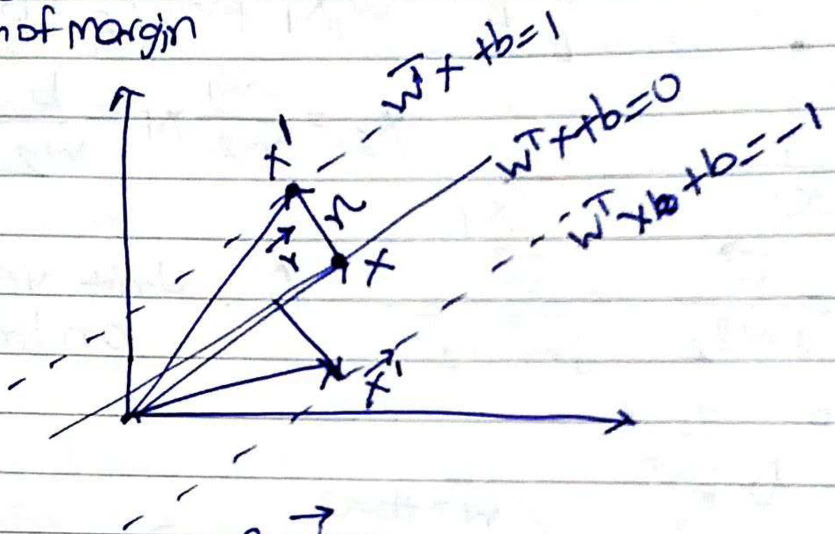
$w^T x + b > 0 \quad y = +1$

$w^T x + b < 0 \quad y = -1$

$w^T x + b = 0$

$\frac{2}{||w||}$

Size or
Length of margin



$$\vec{r} = \frac{r\vec{w}}{||w||}$$

$$\vec{r} = \vec{x'} - \vec{x}$$

$$w^T \vec{r} = w^T x' - w^T x$$

$$\frac{w^T (r\vec{w})}{||w||} = (1-b) - (-b) \qquad \vec{r} = \frac{r\vec{w}}{||w||}$$

$$r \frac{w^T w}{||w||} = +1 \qquad \because w^T w = ||w||_2^2$$

$$r ||w|| = 1$$

$$r = \frac{1}{||w||}$$ Same for other side

$\therefore$ Size of margin $= 2r = \frac{2}{||w||}$

1) We want to learn w, b such that for all point x_i for which $y_i = +1$ $w^T x_i + b > 0$

2) For all point x_i for which $y_i = -1$ $w^T x_i + b < 0$

3) And $\frac{2}{||w||}$ is as large as possible means to maximize the margin.

* Margin is the area around decision boundary in which there is no training data so we not only want line but we want gap between the points so generalization is good.

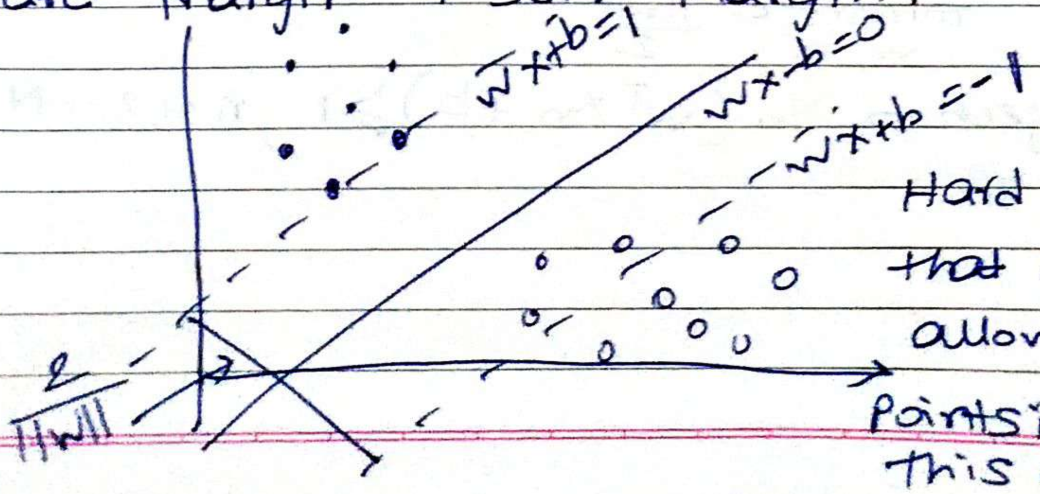
* SVM Learning.

→ IP: Training data - $\{(x_1, y_1), (x_2, y_2) \dots (x_N, y_N)\}$

Objective :- Learn w & b that maximizes the margin.

1) SVM learning task as an optimization problem.
2) Find w & b that gives zero training error.
3) maximizes the margin $[(\frac{2}{||w||})]$
4) Same as minimizing $||w||$

* Hard margin & Soft Margin.



Hard margin that don't allow any points inside this.

* Optimization Formulation

$$\boxed{\begin{array}{l} \underset{w,\,b}{\text{minimize}} \ \frac{\|w\|^2}{2} \\ \text{Subject to } \ y_n(w^T x_n + b) \geq 1, \ n = 1, \cdots N \leftarrow \text{constraint} \end{array}}$$

Optimization with N linear inequality constraint

In this we don't want any point on other side of margin not the inside margin.

$y_n(w^T x_n + b) \geq 1$ is constraint with minimize $\frac{\|w\|^2}{2}$

So SVM is constraint problem with optimization

* Occam's Razor

→ Large margin ⇒ small $\|w\|$
→ Small $\|w\|$ ⇒ Regularized simple solution.
→ Simple Solution ⇒ Better Generalization

* Lagrangian Multipliers for SVM

Optimization Formulation.

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

Subject to $y_n(w^T x_n + b) \geq 1, \ n = 1, 2 \cdots N$

$$\underset{w,b,\alpha}{\text{minimize}}\ L_p(w,b,\alpha) = \frac{\|w\|^2}{2} + \sum_{n=1}^{N} \alpha_n \left(1 - y_n (w^T x_n + b)\right)$$

Subject to $\alpha_n \geq 0 \quad n = 1 \ldots N \quad \forall n$ — (1)

$$\boxed{\min_{w,b}\left(\max_{\alpha} L(w,b,\alpha)\right) \text{ s.t } \alpha \geq 0} \qquad \boxed{\max_{\alpha}\left(\min_{w,b} L(w,b,\alpha)\right)}$$

$$\therefore w = \sum_{n=1}^{N} \alpha^* y_n x_n$$

Only train the example that lie on the margin are relevent these are called support vectors.

For $\alpha_n^* > 0 \quad y_n \{w^T x_n + b\} - 1 = 0$

$$\Rightarrow b = \frac{1}{y_n} - w^T x_n$$

* SVM Testing

For a test point x^* $\rightarrow \alpha_n \neq 0$ only for x_n on margin

$$V = w^T x^* + b$$

$$= \left(\sum \alpha_n^* y_n x_n\right)^T x^* + b$$

$$= \sum_{n=1}^{N} \alpha_n^* y_n x_n^T x_n^*) + b$$

$$= \sum_{\alpha_n > 0} \alpha_n^* y_n (x_n^T x_n^*) + b$$

Only true for support vector.