1/0

Q Derivation

Let 'x' be one of the vectors in the dataset.

Let 'H' be a hyperplane defined through 2 parameter 'w' and 'b'.

Then,

H = set of points ~~which~~ which lie on hyperplane $\{w, b\}$

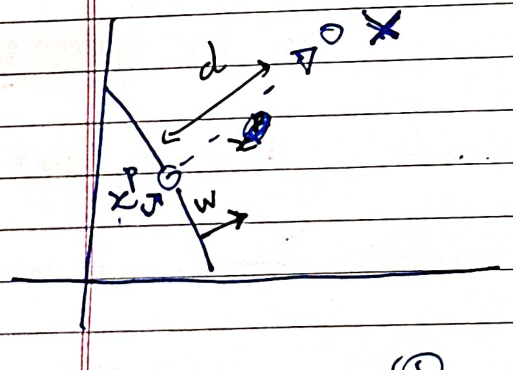
$$= \{x \mid w^T x + b = 0\}.$$

Assuming binary classification right now, let margin 'γ' be the distance from hyperplane to closest point across both classes.

Calculating distance of 'x' to hyperplane 'H'

x = sample point

d = vector from H to x of minimum length.



$$x^P = x - d \quad \text{— (1)}$$

Since d is parallel to w by defn,

let $d = \alpha w$ for some $\alpha \in$ Real.

Since $x_p \in H \Rightarrow w^T x^P + b = 0$ — (2)

Page No.

$|w|$ = length of w

putting ① in ②, we get

$$w^T(x-d)+b = w^T(x-\alpha w)+b=0$$

$$\Rightarrow \alpha = \frac{w^Tx+b}{w^Tw} \quad -③$$

Now, length of 'd' is

$$\|d\| = \sqrt{d^Td} = \sqrt{\alpha^2 w^Tw} \quad \text{using ③}$$

$$= \frac{|w^Tx+b|}{\sqrt{\|w\|^2}} = \frac{|w^Tx+b|}{|w|} \quad -④$$

By defn of margin,

margin of H for a dataset D:

$$\gamma(w,b) = \min_{x\in D} \left(\frac{|w^Tx+b|}{|w|} \right)$$

~~Now, by defn, margin~~

Also, 'γ' ~~to be~~ must be distance H to closest point within both classes.

If not, then we could always move the hyperplane towards the data points of the class that is further away and increase (γ) which is a contradiction to fact that 'γ' was maximized.

H2

Saat

Date ___/___/___

Search for optimal hyperplane

finding max margin seperating hyperplane can be modelled as a constrained optimization problem.
The objective is to MAXIMIZE margin.
constraint: all points must lie on correct side of hyperplane

y_i = label = {1 or -1}

$\therefore$

$$\max_{\{w, b\}} \left(\gamma(w, b)\right) \text{ such that}$$

$$\underbrace{\forall i \rightarrow y_i(w^T x_i + b) \geq 0}_{\text{As}}$$

for $y_i = -1$, $w^T x_i + b < 0$

for $y_i = 1$, $w^T x_i + b > 0$

combine

$\Rightarrow$

$$\max_{w, b} \frac{1}{|w|} \left[\min_{x_i \in D} \left(|w^T x + b|\right)\right] \text{ such that} \quad \text{--- (4)}$$

$$\forall i \rightarrow y_i(w^T x_i + b) \geq 0$$

Since hyperplane is scale invariant i.e, --- (5)

$\gamma(tw, tb) = \gamma(w, b)$ where $t \in R$, $t \neq 0$

(direction for (tw, tb) is same as (w, b))

$\therefore$

Page No.

Bcoz of 5,

we can say that if the optimal hyperplane (w, b) exists such that

$$\min_{x_i \in D} \left(\frac{|w^T x_i + b|}{|w|} \right) = c \quad \text{with constraints satisfied,}$$

then hyperplane $\{\frac{w}{c}, \frac{b}{c}\}$ will also be optimal, where $c > 0$

as

$$\gamma\left(\frac{w}{c}, \frac{b}{c}\right) = \gamma(w, b)$$

with $\min_{x_i \in D} \left|\frac{w^T x_i}{c} + \frac{b}{c}\right| = 1$

and since $y_i (w x_i + b) \geq 0$

would also mean

$$\frac{y_i (w x_i + b)}{c} \geq 0 \quad (\text{for } c > 0) \qquad (6)$$

Due to 6, without loss of generality, let

$$\min_{x_i \in D} \left(w^T x_i + b \right) = 1 \qquad (7)$$

margin now is

$$\gamma(w, b) = \frac{\min_{x_i \in D} |w^T x_i + b|}{|w|} = \frac{1}{|w|}$$

$\Rightarrow$ we need to ~~minimize~~ maximize $\frac{1}{|w|}$ while satisfying constraints

=

$\Rightarrow$ ~~maximize~~ minimize $|w|^2$ while satisfy constraints

Date ___/___/_____

$\therefore$

Problem Reformulated as

Find (w, b) such that

2 constraints

minimize $w^T w$ such that $\begin{cases} \forall i, (w^T x_i + b) y_i \geq 0 \\ \text{AND} \\ \min(|w^T x_i + b|) = 1 \end{cases}$
(w, b)

Now,

$\left\{ \begin{array}{l} \forall i, (w^T x_i + b) y_i \geq 0 \\ \text{AND} \\ \min(|w^T x_i + b|) = 1 \end{array} \right\}$ is same as $\left\{ \begin{array}{c} \forall i \\ y_i (w^T x_i + b) \geq 1 \end{array} \right\}$

$\therefore$ Final formulation

$$\left\{ \min_{w, b} \; w^T w \xrightarrow{} \text{quadratic} \right\}$$

such that $\underbrace{y_i (w^T x_i + b) \geq 1 \quad \forall i}_{\text{constraint} \rightarrow \text{linear in nature}}$

$\therefore$ objective = quadratic
$\therefore$ constraint = linear

$\Rightarrow$ can be solved using QCQP.

Also, $y_i (w^T x_i + b) = 1$ would hold for atleast one vector of each class [called support vectors].

Page No.

45

Date ___/___/___

∴ Problem now reduced to

Find the simplest hyperplane (i.e, one with minimum $w^T w$)

such that all input samples lie atleast 1 unit away from hyperplane on correct side.

Hence, proved.