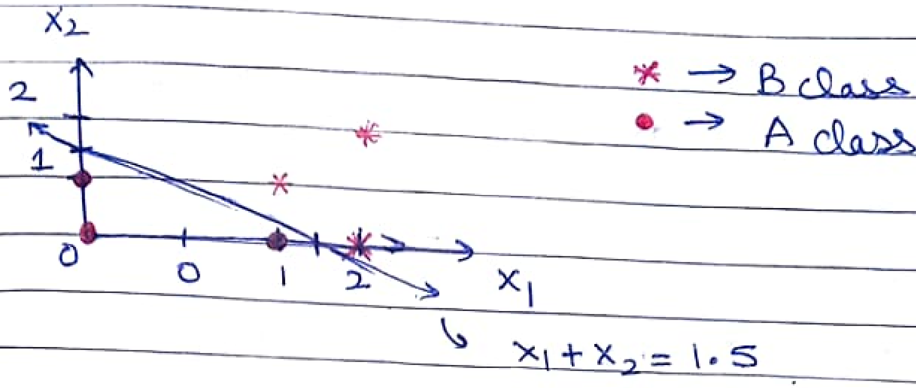
Page No.
Date :

ML - Assignment

Name - Ishit Bajpai
Roll. No - 2020380

1. Plotting the points



Yes, points are linearly separable as we can draw line $x_1 + x_2 = 1.5$ separating both the classes completely
~~Now Showing~~
for Class A
$(x_1 + x_2 - 1.5)$

x_1	x_2	$x_1+x_2-1.5$
0	0	-1.5
1	0	-0.5
0	1	-0.5

$x_1 + x_2 - 1.5 < 0$, for all points in Class A $\Rightarrow$
All points in Class A lies below $x_1 + x_2 = 1.5$ (1)
Similarly, for class B

x_1	x_2	$x_1+x_2-1.5$
1	1	0.5
2	2	2.5
2	0	0.5

$x_1 + x_2 - 1.5 > 0$, for all points in Class B
$\Rightarrow$ All points in Class B lies above $x_1 + x_2 = 1.5$ (2)

Page No.
Date :

from ① and ② linearly separa
we can conclude that line ~~divides~~
Class A and B.

2. As discussed in the lectures.
Our primal form $\min \frac{1}{2}\|w\|^2$ st

$y_i(w^T x_i + b) \geq 1$ was converted

to dual where we are required to

maximise $Q(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$

~~st~~ such that $\alpha_i \geq 0, \forall i = 0, \dots n.$
and $\sum_{i=1}^{n} \alpha_i y_i = 0$

(n is the number of traing examples)

So, our optimization problem is

maximise $Q(\alpha)$ st $\alpha_i \geq 0 \ \forall i = 0, 1 \dots N$
, $\sum_{i=1}^{n} \alpha_i y_i = 0.$

Since Optimal hyperplane has to be in between the points of Class A and Class B.
Therefore ~~Only~~ ~~Supporting~~ Potential Supporting Vectors can be ~~x1~~ (1, 0) and ~~x2~~ (0, 1)
from Class A

Page No.
Date :

(0,0) can not be a supporting vector as it lies below (1,0) in terms of x_1 and (0,1) in terms of x_2.

Similarly potential supporting vectors from class B can be (1,1) and (2,0)

(2,2) can not be a supporting vectors as it lies ~~below~~ above (1,1) in terms of x_1 and (2,0) in terms x_2.

Page No.
Date :

Let us consider 3 SV to form optimal hyperplane.

$S_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $S_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ $S_3 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

We'll add one bias to each of the vector (augumenting) (Later this will store our bias)

$\tilde{S}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $\tilde{S}_2 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}$ $\tilde{S}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Using Equation

$\sum_{i=1}^{N} \alpha_i \tilde{S}_i \tilde{S}_j =$ Sign of Class to which $\tilde{S}_j$ belongs.

dot product

α_i's → Lagranganian multipliers

for S_1

$\alpha_1 \tilde{S}_1 \tilde{S}_1 + \alpha_2 \tilde{S}_1 \tilde{S}_2 + \alpha_3 \tilde{S}_1 \tilde{S}_3 = +1$

$\alpha_1 (1+1+1) + \alpha_2 (2+0+1) + \alpha_3 (1+0+1) = +1$

$3\alpha_1 + 3\alpha_2 + 2\alpha_3 = +1$ — (1)

for S_2 $\alpha_1 \tilde{S}_1 \tilde{S}_2 + \alpha_2 \tilde{S}_2 \tilde{S}_2 + \alpha_3 \tilde{S}_3 \tilde{S}_2 = +1$

$\alpha_1 (2+0+1) + \alpha_2 (4+0+1) + \alpha_3 (2+0+1) = 1$

$3\alpha_1 + 5\alpha_2 + 3\alpha_3 = 1$ — (2)

Page No.
Date :

~ for S_3

$$\alpha_1 \tilde{s}_1 \tilde{s}_3 + \alpha_2 \tilde{s}_2 \tilde{s}_3 + \alpha_3 \tilde{s}_3 \tilde{s}_3 = -1$$

$$2\alpha_1 + 3\alpha_3 + 2\alpha_3 = -1 \quad \text{--- (3)}$$

Solving $\alpha_1, \alpha_2, \alpha_3$, we get

$$\alpha_1 = 2, \ \alpha_2 = 5, \ \alpha_3 = -10$$

$$\hat{\omega} = \sum_{i=1}^{n} \alpha_i \tilde{s}_i$$

$$= 2\begin{pmatrix}1\\1\\1\end{pmatrix} + 5\begin{pmatrix}2\\0\\1\end{pmatrix} - 10\begin{pmatrix}1\\0\\1\end{pmatrix}$$

$$= \begin{pmatrix}2+10-10\\2+0-0\\2+5-10\end{pmatrix} = \begin{pmatrix}2\\2\\-3\end{pmatrix}$$

from previous argument

$$w = \begin{pmatrix}2\\2\end{pmatrix} \qquad \text{bias} = -3$$

Eqⁿ of ~~optimal~~ hyperplane (maximum margin)

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

$$\begin{bmatrix}2 & 2\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} - 3 = 0 \Rightarrow 2x_1 + 2x_2 = 3$$

$$\Rightarrow \boxed{x_1 + x_2 = 1.5}$$

$x_1+x_2 = 1.5$ Page No.
Date :

Calculating distance of points from hyp

~~(0,0)~~

Class A

(i) $(0,0)$ $\quad \dfrac{|0+0-1.5|}{\sqrt{2}} = \dfrac{3}{2\sqrt{2}}$

S_3

(ii) $(1,0)$ $\quad \dfrac{|1+0-1.5|}{\sqrt{2}} = \dfrac{1}{2\sqrt{2}}$

(iii) $(0,1)$ $\quad = \dfrac{|0+1-1.5|}{\sqrt{2}} = \dfrac{1}{2\sqrt{2}}$

- Since it is at same distance from $x_1+x_2 = 1.5$ as $S_3 \Rightarrow (0,1)$ is also a Support Vector.

Let $S_4 = (0,1)^T$

Class B

S_1

(i) $(1,1) \rightarrow \dfrac{|1+1-1.5|}{\sqrt{2}} = \dfrac{1}{2\sqrt{2}}$

(ii) $S_2\ (2,0) \rightarrow \dfrac{|2+0-1.5|}{\sqrt{2}} = \dfrac{1}{2\sqrt{2}}$

(iii) $(2,2) \rightarrow \dfrac{|2+2-1.5|}{\sqrt{2}} = \dfrac{5}{2\sqrt{2}} =$

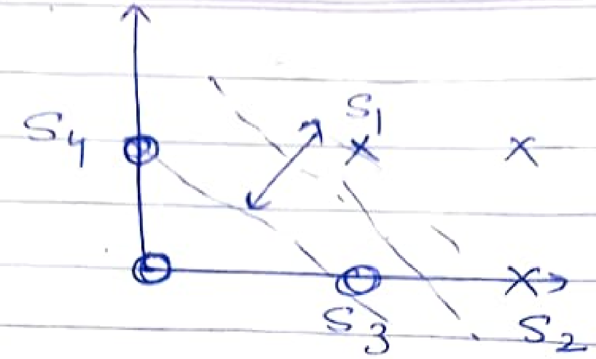
Support Vectors are

$(1,0), (0,1), (1,1), (2,0)$

(3)

(margin)

$M = \frac{1}{\sqrt{2}}$



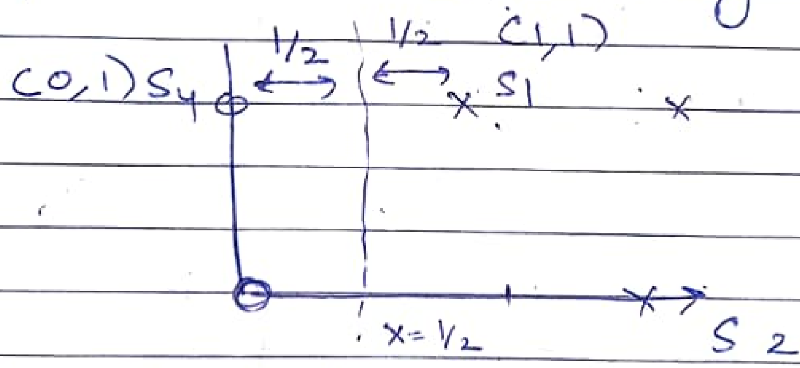
$S_1 = (1, 1)^T \quad S_2 = [2 \ 0]^T \quad S_3 = [1 \ 0]^T$

$S_4 = [0 \ 1]^T$

★ Since we constructed optimal margin using S_1, S_2, S_3, ∴ removing S_4 will have no effect on optimal margin. $\left(\frac{1}{2\sqrt{2}} \times 2\right)$

(discussed previous)

• if we remove S_3, ~~then by~~



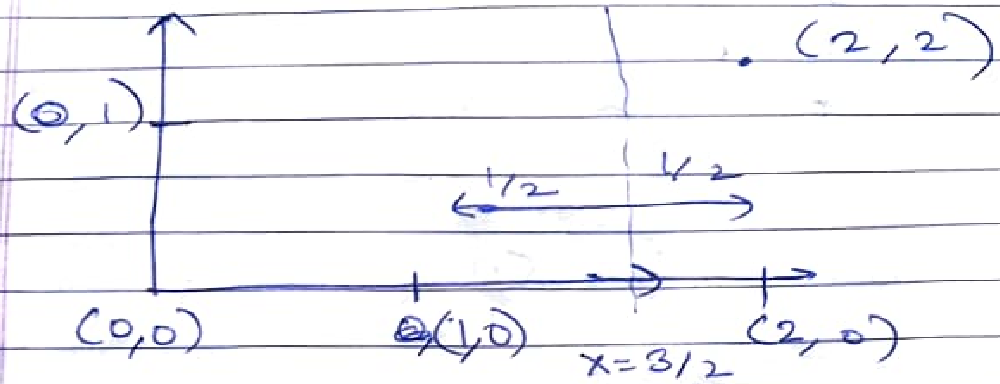
by visual inspection we can see that S_2 no longer remains a support vector and (0,0), (0,1), (1,1) becames the new support vectors

The new optimal hyperplane became $x = \frac{1}{2}$. (all support vectors are at equal distance from $x = 1/2$)

★ new margin = 1, Thus margin ↑ if we remove (1,0).

Page No.
Date :

if we remove S_1 (1,1)



(2,2), (1,0), (2,0) ~~are~~ $\tilde{S}$ becomes the new support Vector
and x=1.5 is optimal hyperplane as it is at equal distance from all the support vectors.
new margin becomes 1
⇒ Margin ↑ if we remove (1,1).

- Removing S_2 (2,0).

$S_1(1,1)$ $\quad S_2(2,0)$ $\quad S_4(0,1)$.

Using $\sum \alpha_i \tilde{s}_i \tilde{S}_j$ = Sign of Class to which S_j belongs

$$\alpha_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} + \alpha_4 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1$$

$$3\alpha_1 + 3\alpha_2 + 2\alpha_4 = 1 \quad — ①$$

Page No.
Date :

$$\alpha_1 \tilde{S}_1 \tilde{S}_2 + \alpha_2 \tilde{S}_2 \tilde{S}_2 + \alpha_4 \tilde{S}_4 \tilde{S}_2 = +1$$

$$3\alpha_1 + 5\alpha_2 + \alpha_4 = 1 \quad — (2)$$

$$\alpha_1 \tilde{S}_1 \tilde{S}_4 + \alpha_2 \tilde{S}_2 \tilde{S}_4 + \alpha_4 \tilde{S}_4 \tilde{S}_4 = -1$$

$$2\alpha_1 + \alpha_2 + 2\alpha_4 = -1 \quad — (3)$$

$$\alpha_4 = -10, \alpha_2 = -5, \alpha_1 = 12$$

$$\tilde{\omega} = \sum \alpha_i \tilde{S}_i$$

$$\tilde{\omega} = 12\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} - 5\begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} - 10\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 12-10 \\ 12-10 \\ 12-5-10 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ -3 \end{pmatrix}$$

$$\omega = [2 \quad 2]^T \qquad \text{bias} = -3$$

$$\omega^T x + b = 0 \Rightarrow 2x_1 + 2x_2 - 3 = 0$$

Thus we get the Same optimal hyperplane as we would get with S_1, S_2, S_3, S_4.

$\Rightarrow$ On removing $(2,0)$ margin remains same.

Page No.
Date :

Q4 ~~In~~ SVM, we are maximising the margin. When

*

(Q4) In general, for any dataset if we remove Support Vectors then optimal margin can increase or remain same (as shown in prev questions ~~it~~ it remains same).

~~Ans~~

- We can expect an increase because on removing support vector distance between the nearest points of opposite classes can increase, as a result margin can also increase (as optimal hyperplane is made st $d^+ = d^-$) * (Illustration shown at last)

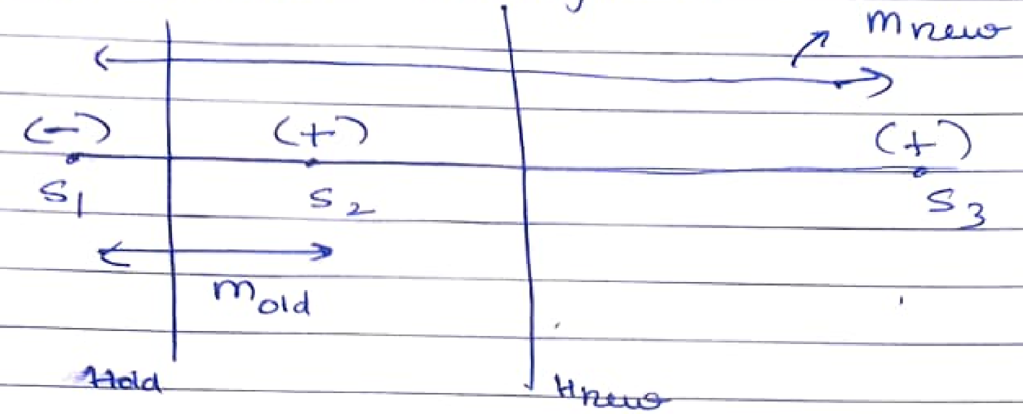
However it is not ~~nec~~ necessary that margin will ↑ as in previous question even on dropping S_4, S_2 margin remained same.

Mathematical Approach: SVM is maximisation problem in which constraints are made because of support vectors

On dropping support vectors, in a way constraints are somewhat relaxed. This may result in more maximization than before.

* As

Case when margin ↑.



Let H_{old} is maximum margin hyperplane between (−) and (+).
M_{old} is the margin length, Suppose we now drop S_2 ~~and~~, S_3 is now closest point of + class to − (S_1).
New optimal hyperplane becomes H_{new} and very clearly $m_{new} > m_{old}$.

Thus margin ↑.