Name - Anushree Jumade (axj174430)

Machine Learning Homework IV

Q.1.

X_1	X_2	Y
-1	1	+
1	-1	−

To find :- Linear SVM classifier.

→ 2 datapts given

∴ 2 Lagragian multipliers. → One for each pt.

Dual formula :-

$$L = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j$$

	P_1	P_2
P_1	2	-2
P_2	-2	2

$P_1 \cdot P_1$ $\quad \frac{-1 \cdot -1}{+\ 1 \cdot 1} = +\frac{1}{1} = 2$

$P_1 \cdot P_2$ $\quad -1 \cdot (1) + (1)(-1) = -2 = P_2 \cdot P_1$

$P_2 \cdot P_2 = 1 \cdot 1 + (-1)\cdot(-1) = 2$

Substituting the dot products & y-values we get

$$L = \alpha_1 + \alpha_2 - \frac{1}{2}\alpha_1^2(1)(2) - \frac{1}{2}\alpha_1\alpha_2(-1)(-2) - \frac{1}{2}\alpha_1 \cdot \alpha_2(-1)(-2) - \frac{1}{2}\alpha_2^2(1)(2)$$

$$L = \alpha_1 + \alpha_2 - \alpha_1^2 - 2\alpha_1\alpha_2 - \alpha_2^2 \quad \text{— (I)}$$

Differentiating L w.r.t w & b.

$$w = \sum_i \alpha_i y_i x_i$$

$$\sum \alpha_i y_i = 0$$

$$\alpha_1 - \alpha_2 = 0 \rightarrow \alpha_1 = \alpha_2$$

Differentiating L w.r.t α_1 & α_2, we get

$$1 - 2\alpha_1 - 2\alpha_2 = 0 \qquad 1 - 2\alpha_1 - 2\alpha_2 = 0$$

$\alpha_1 + \alpha_2 = \frac{1}{2}$.

& $\alpha_1 = \alpha_2$

$\therefore \alpha_1 = \alpha_2 = \frac{1}{4}$.

By substituting the α_1 & α_2 in w eqⁿ, we get

$W = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2$.

$$W = \begin{matrix} \frac{1}{4} \times (1)(-1) + \frac{1}{4}(-1)(1) \\ \frac{1}{4} \times (+1)(1) + \frac{1}{4}(-1)(-1) \end{matrix} = \begin{matrix} -1/2 \\ 1/2 \end{matrix}.$$

Substitute w in any one of the initial conditions

$w \cdot x + b = +1 \longrightarrow$ +ve SV.

$w \cdot x + b = -1 \longrightarrow$ -ve SV.

we get $\boxed{b = 0}$

Ans $\rightarrow$ $\alpha_1 = \alpha_2 = 1/4$

$w_1 = -\frac{1}{2}$ $w_2 = \frac{1}{2}$.

$b = 0$

Q3

	X_1	X_2	class
1	-1	-1	-1
2	1	1	1
3	0	2	1

α_1, α_2 & α_3 be the Lagrangian multipliers for the 3 data pts

a. $K(x_i, x_j) = (1 + x_i^T x_j)^d$ x_i & x_j are I/P vectors

We maximize the f^n :.

optimization problem to solve

$$\sum_{i=1}^{3} \alpha_i - \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j \left(1 + x_i^T x_j\right)^2 .$$

$\alpha_1, \alpha_2, \alpha_3 \geq 0$ & $-\alpha_1 + \alpha_2 + \alpha_3 = 0$. $(\sum \alpha_i y_i = 0)$.

The quantity

$y_i y_j \left(1 + x_i^T x_j\right)^2$ for different values of i & j can be computed as :.

$y_1 = -1$ $\quad x_1 = -1, -1$

$y_2 = 1$ $\quad x_2 = 1, 1$

$y_3 = 1$ $\quad x_3 = 0, 2$.

$i = 1 \quad j = 1$ $\quad y_1 \cdot y_1 = -1 \times -1 = 1$.

$$x_1^T \cdot x_1 = [-1, -1] \begin{bmatrix} -1 \\ -1 \end{bmatrix} = 2 .$$

$$\therefore y_i y_j \left(1 + x_i^T x_j\right)^2 = 1 \left(1 + 2\right)^2 = 9 .$$

Similarly.

$i = 1, j = 2$ $\quad y_1 \cdot y_2 = -1$.

$$x_1^T \cdot x_2 = \begin{bmatrix} -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = -2 .$$

$$\therefore y_i y_j \left(1 + x_i^T x_j\right)^2 = -1 \left(1 + (-2)\right)^2 = -1 .$$

Thus, for different values of i & j, we get a matrix of value

$$\begin{pmatrix} 9 & -1 & -1 \\ -1 & 9 & 9 \\ -1 & 9 & 25 \end{pmatrix}$$

For the condition $\alpha_1 = \alpha_2 = 1/8$ & $\alpha_3 = 0$. & $b = 0$.

b) The support vectors are pts 1 & 2 because.
α_1 & α_2 > 0 .

c). The dot product of the kernel with the test point is $(4, 0)$

$$\therefore -\alpha_1 \cdot 4 + \alpha_2 \cdot 0 \; = \; \frac{-1}{8}(4) + \frac{1}{8}(0) < 0$$

Therefore, the class is -1.

d). $\alpha_1 = \alpha_2 = \frac{1}{8}$ $\alpha_3 = 0$ & $b = 0$.
The dot product of the kernel with the test pt is $(0, 4)$.

$$-\frac{1}{8}(0) + \frac{1}{8}(4) > 0$$

Therefore, the class is $+1$.

Q. 4.

X	Y
-2	1
-1	-1
1	-1
2	1

optimization problem to solve.

Dual formula :-

$$L = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j \, y_i y_j \, x_i x_j$$

$y_1 = 1$ $y_2 = -1$ $y_3 = -1$ $y_4 = 1$.
$x_1 = -2$ $x_2 = -1$ $x_3 = 1$ $x_4 = 2$.

For the 4x4 matrix ; we can proceed as :-

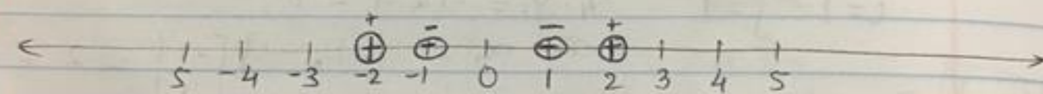
$y_i \cdot y_j \, x_i^T x_j \rightarrow (+1)(+1)(2 \cdot 2) = 4$. $i = j = 1$
$\rightarrow (1)(-1)(-2)(-1) = -2$ $i = 1, j = 2$.

& so on.

The quantity $y_i y_j x_i^T x_j$ for different values of i & j is given by

$$\begin{pmatrix} 4 & -2 & 2 & -4 \\ -2 & 1 & -1 & 2 \\ 2 & -1 & 1 & -2 \\ 4 & 2 & -2 & 4 \end{pmatrix}$$

b) No, zero training error on this dataset not possible because the data is not linearly separable.



c) maximise $\sum_{i=1}^{4} \alpha_i - \frac{1}{4} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j (1 + x_i^T x_j)^2$

subject to constraint :- $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \geq 0.$

$\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 0 \qquad \sum_{i=1}^{4} \alpha_i y_i = 0$

The quantity $y_i y_j (1 + x_i^T x_j)^2$ for different values of i & j is given by the cells

$$\begin{pmatrix} 25 & -9 & -1 & 9 \\ -9 & 4 & 0 & -1 \\ -1 & 0 & 4 & -9 \\ 9 & -1 & -9 & 25 \end{pmatrix}$$

For ex :

$i=1 \; j=1. \rightarrow 1 \times 1 \times (1 + (-2)(-2))^2$

$= 1(5)^2 = 25.$

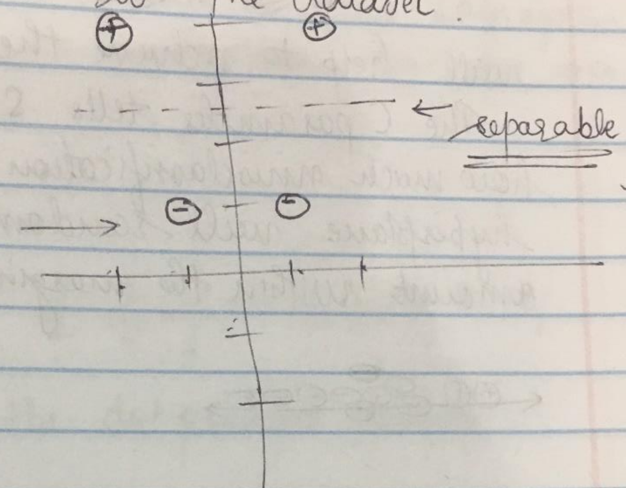
$i=2, j=1. \quad -1 \times 1 (1 + (-1)(-2))^2$

$= -1(1+2)^2 = -9$

$i=2 \; j=3 \rightarrow (-1)(-1)[1 + (-1)(1)] = 0.$

d) yes, the quadratic kernel will correctly classify this data because it will add the feature x^2 to the dataset.

x	x^2	y
-2	4	1
-1	1	-1
1	1	-1
2	4	1



Q.5 The given statement is FALSE.
Support Vector Machine has a scoring function which computes a 'score' for a new input. SVM is a binary classifier; if the o/p of the scoring function is -ve, then the i/p is classified as belonging to class $y=-1$. If the score is positive, the i/p is classified as belonging to class $y=1$.
The score value isn't the probability value.

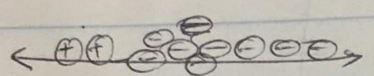
Q 6.

6. Adding the constraint.

$$0 \leq \alpha_i \leq C$$ for negative pts only

will help to achieve the said goal!

The C parameter tells SVM optimization how much misclassification to avoid. The optimal hyperplane will randomly shift by a small amount within the margin.



~~Q 7.~~

Q.2

X_1	X_2	Y
1	0	+
-1	2	−
0	-1	+

Since there are two pts, there will be two Lagragian multipliers, one for each pt

Dual formula :

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i x_j .$$

Calculating the dot product matrix .

	P_1	P_2	P_3
P_1	1	-1	0
P_2	-1	5	-2
P_3	0	-2	1

Substituting the dot products & y-values we get .

$$L = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2 (1)(1) - \frac{1}{2}\alpha_1\alpha_2(-1)(-1)$$
$$-\frac{1}{2}\alpha_1\alpha_3(1)(0) - \frac{1}{2}\alpha_1\alpha_2(-1)(-1)$$
$$-\frac{1}{2}\alpha_2^2(1)(5) - \frac{1}{2}\alpha_2\alpha_3(-1)(-2)$$
$$-\frac{1}{2}\alpha_3\alpha_1(1)(0) - \frac{1}{2}\alpha_3\alpha_2(-1)(-2)$$
$$-\frac{1}{2}\alpha_3^2(1)(1).$$

$$L = \alpha_1 + \alpha_2 + \alpha_3 - \frac{1}{2}\alpha_1^2 - \alpha_1\alpha_2 - \frac{5}{2}\alpha_2^2 - 2\alpha_2\alpha_3 - \frac{1}{2}\alpha_3^2 \quad — (2)$$

When we differentiate L w.r.t w & b.

$$w = \sum_i \alpha_i y_i x_i .$$

$$\sum_i \alpha_i y_i = 0 \longrightarrow \alpha_1 - \alpha_2 + \alpha_3 = 0 \quad — (1).$$

Differentiate (2) w.r.t α_1, α_2 & α_3 we get the following three eq^ns -

$$1 - \alpha_1 - \alpha_2 = 0 . \quad — I$$

$$1 - \alpha_1 - 5\alpha_2 - 2\alpha_3 = 0 . \quad — II$$

$$1 - 2\alpha_2 - \alpha_3 = 0 \quad — III$$

from (1) $\rightarrow$ $\alpha_2 = \alpha_1 + \alpha_3$.

From II $\rightarrow$. $\frac{1 - \alpha_1 - 2\alpha_3}{5} = \alpha_2$.

$$5(\alpha_1 + \alpha_3) = 1 - \alpha_1 - 2\alpha_3 .$$

$$6\alpha_1 = 1 - 7\alpha_3 .$$

I & III :

$$\alpha_2 = 1 - \alpha_1$$

$$\alpha_2 = \frac{1 - \alpha_3}{2} .$$

$$2(1 - \alpha_1) = (1 - \alpha_3)$$

$$2 - 2\alpha_1 = 1 - \alpha_3$$

$$3 + 3\alpha_3 = 6\alpha_1 .$$

$\alpha_1 = \frac{1}{2}, \quad \alpha_2 = \frac{1}{2} \quad \alpha_3 = 0.$

Substituting values of α_1 & α_2 in w, we get.

$$w = \begin{matrix} \frac{1}{2}(+1)(1) + \frac{1}{2}(-1)(-1) \\ \frac{1}{2}(+1)(0) + \frac{1}{2}(-1)(2) \end{matrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

substitute w in any one of the initial conditions :

$w.x + b \Rightarrow +1 \longrightarrow$ Positive SV

$wx + b = -1 \longrightarrow$ Negative.

we get $b = 0$ & $b = 2$

so avg ~~is~~ value of $\boxed{b = 1}$.

Ans $\rightarrow \alpha_1 = \frac{1}{2} \quad \alpha_2 = \frac{1}{2} \quad \alpha_3 = 0.$

$w = [1, -1]$

$b = 1.$