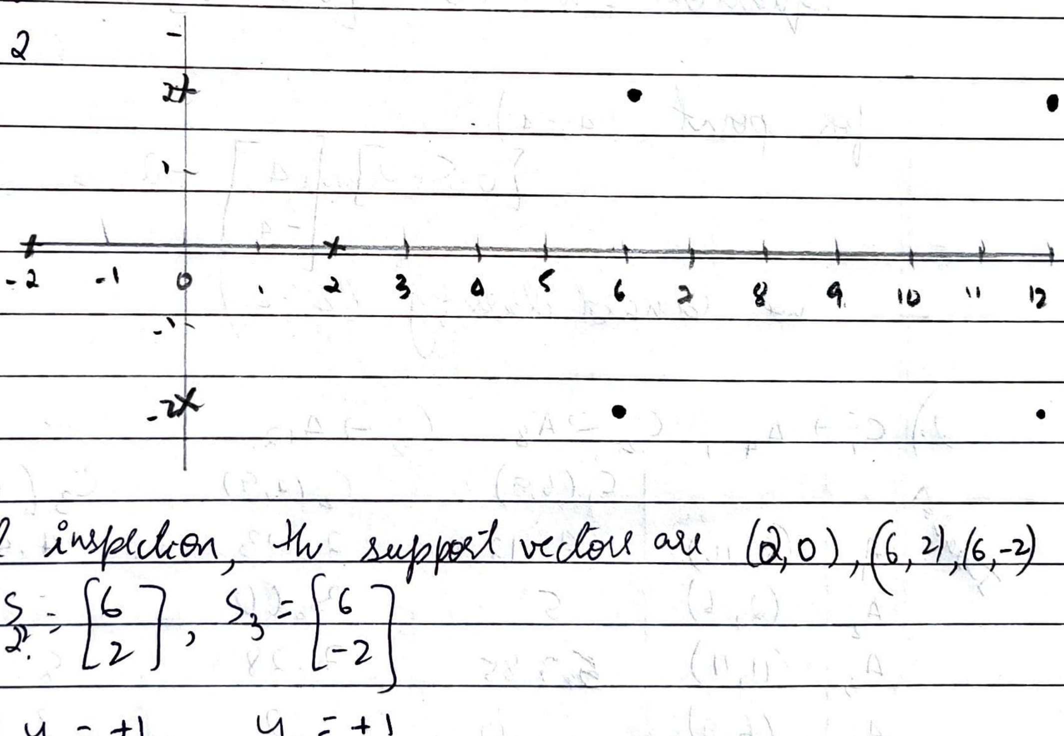
Date______ Page______

Machine Intelligence
Assignment 2.

Find a SVM Classifier for the given data and classify (4, -4)

x_1	x_2	label
6	-2	+
12	2	+
6	2	+
12	-2	+
0	-2	-
-2	0	-
2	0	-
0	2	-



By manual inspection, the support vectors are (2, 0), (6, 2), (6, -2)

$S_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ $S_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$, $S_3 = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$

$y_1 = -1$ $y_2 = +1$ $y_3 = +1$

using the binding constraint equations:-

$w^T x + b = -1$, $w^T x + b = 1$ $x \rightarrow$ support vector.

we get

$w^T S_1 + b = -1$ $w^T S_3 + b = +1$

$w^T S_2 + b = +1$

substituting.

$$2w_1 + b = -1$$
$$6w_1 + 2w_2 + b = \phi$$
$$6w_1 - 2w_2 + b = 1$$

solving above

$$w_1 = 0.5$$
$$w_2 = 0$$
$$\cancel{w_2}\ b = -2$$

$\therefore \quad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix} \qquad$ bias $= -2$.

equation $= w^T x + b = \begin{bmatrix} 0.5 & 0 \end{bmatrix} x - 2$

for point $(4, -4)$

$$\begin{bmatrix} 0.5 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ -4 \end{bmatrix} - 2 = 2 - 2 = 0.$$

we cannot classify $(4, -4)$

2) $C_1 \to A_4$, $C_2 \to A_8$, $C_3 \to A_{12}$.

Iteration)

		$C_1(6,9)$	$C_2(4,9)$	$C_3(4,6)$	
A_1	(2, 10)	4.123	2.23	4.472	C_2
A_2	(2, 6)	5	3.60	2.	C_3
A_3	(11, 11)	5.385	7.28	8.6	C_1
A_4	(6, 9)	0	2.	3.6	C_1
A_5	(6, 4)	5	5.3	2.8	C_3
A_6	(1, 2)	8.6	7.6.	5	C_3
A_7	(5, 10)	1.41	1.41	4.12	C_2
A_8	(4, 9)	2	0	3.	C_2
A_9	(10, 12)	5	6.706	8.48	C_1

Date ______ Page ______

		(6,9)	(4,9)	(4,6)	
A_{10}	(7,5)	4.12	5	3.16	C_3
A_{11}	(9,11)	3.6	5.38	7.07	C_1
A_{12}	(4,6)	3.6	3	0	C_3
A_{13}	(3,10)	3.16	1.41	4.12	C_2
A_{14}	(3,8)	3.16	1.41	2.23	C_2
A_{15}	(6,11)	2	2.82	5.38	C_1

$C_1 :- \dfrac{11+6+10+9+6}{5}, \dfrac{11+9+12+11+11}{5}$

C_1 (8.4, 10.8) $C_3 = (4, 4.6)$

$C_2 = (3.4, 9.4)$

		8.4,10.8	3.4,9.4	4,4.6	
A_1	(2,10)	6.44	1.523	5.75	C_2
A_2	(2,6)	8	3.67	2.44	C_3
A_3	(11,11)	2.60	7.76	9.48	C_1
A_4	(6,9)	3	2.63	4.8	C_2
A_5	(6,4)	7.21	5.99	2.08	C_3
A_6	(1,2)	11.49	7.77	3.96	C_3
A_7	(5,10)	3.49	1.7	5.49	C_2
A_8	(4,9)	4.75	6.7	4.4	C_2
A_9	(10,12)	2	7.09	9.52	C_1
A_{10}	(7,5)	5.9	5.68	3.02	C_3
A_{11}	(9,11)	0.6	5.82	8.12	C_1
A_{12}	(4,6)	6.5	3.45	1.4	C_3
A_{13}	(3,10)	5.4	0.72	5.49	C_2
A_{14}	(3,8)	6.08	1.45	3.54	C_2
A_{15}	(6,11)	2.9	3.05	6.76	C_1

Cluster C_1 (9, 11.25)

(11, 11), (10, 12), (9, 11), (6, 11)

&

Cluster C_2 (3.83, 9.33)

(2, 10), (6, 9), (5, 10), (4, 9), (3, 10), (3, 8)

Cluster C_3 (4, 4.6

(2, 6), (6, 4), (1, 2), (7, 5), (4, 6)

3. For SVM, derive

i) The optimization problem that needs to be solved for finding the optimal hyperplane

A Given a linearly separable training set $D = \{(x_i, y_i) \mid x_i \in \mathbb{R}^n, y_i \in \{-1, 1\}\}_{i=1}^{m}$, and a hyperplane with a normal vector w and bias b, geometric margin M

$$M = \min_{i=1 \ldots m} \gamma_i \quad \text{where} \quad \gamma_i = \left(y_i \frac{w \cdot x_i}{\|w\|} + \frac{b}{\|w\|} \right)$$

The optimal separating hyperplane is the hyperplane defined by the normal vector w & bias b for which M is largest.

To find w & b, we need to solve the following optimization problem, with the constraint that the margin of each sample should be greater or equal to M.

$$\underset{w, b}{\text{maximize}} \; M$$

$$\text{subject to} \quad \gamma_i \geq M, \; i = 1, \ldots m$$

~~the sub~~

where $M = \dfrac{F}{\|w\|}$ (relationship b/w geometric and functional margin)

Date ______ Page ______

rewritten $$\underset{w,b}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \frac{f_i}{\|w\|} \geq \frac{F}{\|w\|}, \quad i=1,\dots m.$$

Removing norm on b.s

$$\underset{w,b}{\text{maximize}} \quad M$$

$$\text{subject to} \quad f_i \geq F \quad , i=1,\dots n$$

Since the scale of w & b does not matter, decide the scale w & b so that $F=1$

i.e

$$\underset{w,b.}{\text{maximize}} \quad M$$

$$\text{subject to} \quad f_i \geq 1 \quad , i=1,\dots m.$$

subbing $M = \frac{F}{\|w\|}$

$$\underset{w,b}{\text{maximize}} \quad \frac{F}{\|w\|}$$

$$\text{subject to} \quad f_i \geq 1 \quad , i=1,\dots m$$

$F=1$

$$\underset{w,b}{\text{maximize}} \quad \frac{1}{\|w\|}$$

$$\text{subject to} \quad f_i \geq 1 \quad i=1,\dots n.$$

equivalent to

$$\underset{w,b}{\text{minimize}} \quad \|w\| \qquad f_i = y_i(w \cdot x_i) + b$$

$$\text{subject to} \quad y_i(w \cdot x_i) + b \geq 1 \quad , i=1,\dots n.$$

equivalent to

$$\underset{w,b}{\text{minimize}} \quad \frac{1}{2}\|w\|^2$$

$$\text{subject to} \quad y_i(w \cdot x_i) + b \geq 1 \quad , i=1,\dots m.$$

ii) The Lagrangian for the optimization problem.

SVM optimization problem $\underset{w,b}{\text{minimize}} \frac{1}{2}\|w\|^2$

subject to $y_i(w \cdot x_i + b) - 1 \geq 0$, $i = 1 \ldots n$

objective function to minimize. $f(w) = \frac{1}{2}\|w\|^2$

m constraint functions $g_i(w,b) = y_i(w \cdot x_i + b) - 1$, $i = 1, \ldots n$

Introducing Lagrangian function

$$L(w, b, \alpha) = f(w) - \sum_{i=1}^{m} \alpha_i g_i(w,b)$$

subbing

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{m} \alpha_i \left[y_i(w \cdot x_i + b) - 1\right]$$

to find solution solve. $\underset{w,b}{\min} \; \underset{\alpha}{\max} \; L(w,b,\alpha)$

subject to $\alpha_i \geq 0$, $i = 1, \ldots m$

ii) The dual form of the optimization problem.

Wolfe dual

Solving the Lagrangian primal problem of

$$\underset{w,b}{\min} \; \underset{\alpha}{\max} \; L(w,b,\alpha)$$

subject to $\alpha_i \geq 0$, $i = 1, \ldots n$.

Solving by taking partial derivative of L w.r.t. to w & b.

$$\nabla_w L = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \quad - (1)$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^{n} \alpha_i y_i = 0. \quad - (2)$$

Date________ Page________

from (1) $\quad w = \sum_{i=1}^{m} \alpha_i y_i x_i$

Subbing w in L.

$$W(\alpha, b) = \frac{1}{2}\left(\sum_{i=1}^{n}\alpha_i y_i x_i\right)\left(\sum_{j=1}^{m}\alpha_j y_j x_j\right) - \sum_{i=1}^{n}\alpha_i\left[y_i\left(\left(\sum_{j=1}^{n}\alpha_j y_j x_j\right)x_i + b\right) - 1\right]$$

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i x_j - \sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i \cdot x_j - b\sum_{i=1}^{n}\alpha_i y_i + \sum_{i=1}^{m}\alpha_i$$

$$= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i \cdot x_j - b\sum_{i=1}^{n}\alpha_i y_i$$

since $\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{m}\alpha_i y_i = 0$

$$\therefore \quad W(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j x_i \cdot x_j \quad \rightarrow \text{Wolfe dual Lagrangian function}$$

Wolfe dual problem:

$$\underset{\alpha}{\text{maximize}} \ \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i \cdot x_j$$

subject to $\quad \alpha_i \geq 0$ for any $i = 1, \dots n$.

$$\sum_{i=1}^{n}\alpha_i y_i = 0$$

ii) Equations for w, b in terms of lagrange multipliers

Compute w.

as derived above $\quad w = \sum_{i=1}^{m}\alpha_i y_i x_i \quad$ from $\nabla_w L$

Compute b.

$$y_i(w \cdot x_i + b) - 1 \geq 0.$$

the above equation says that the closest points to the hyperplane will have a functional margin of 1. We chose 1 as the value when we decided to scale w.

$\therefore \quad y_i(w.x_i + b) = 1$

multiply b.s by y_i
$(\because y_i^2 = 1)$ $\quad w.x_i + b = y_i$
$$\boxed{b = y_i - w.x_i}$$

given:-

4) event A = raining. $\quad P(A) = 0.4 \quad P(B|A) = 0.45$
event B = cloudy. $\quad P(B) = 0.6$

Bayestheorm.

Probability that it is raining given it is cloudy $= P(A|B)$

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

$$= \frac{0.45 \times 0.4}{0.6} = 0.3$$

5. given.
2 classes → A, B
2 feature → X, Y.

for A. $\quad$ B.
$P(X=1|A) = 0.2 \quad P(X=1|B) = 0.8$
$P(Y=1|A) = 0.3 \quad P(Y=1|B) = 0.7$
priors:- $P(A) = 0.3 \quad P(B) = 0.7$

find class for instance $X=1, Y=1$ using Bayes Optimal classifier.

$$\underset{v_j \in V}{\text{argmax}} \sum_{h_i \in H} P(v_j|h_i) * P(h_i|D)$$

Date______ Page______

to find

$$P(A|x=1, Y=1) = \frac{P(x=1, Y=1|A)\,P(A)}{P(x=1, Y=1)}$$

$$P(B|x=1, Y=1) = \frac{P(x=1, Y=1|B)\,.\,P(B)}{P(x=1, Y=1)}$$

$$P(A|x=1, y=1) = \frac{P(x=1|A)\,P(Y=1|A).\,P(A)}{P(x=1, Y=1|A).P(A) + P(x=1, Y=1|B).P(B)} = \frac{0.2\times0.3\times0.3}{0.06\times0.3 + 0.56\times0.7}$$

$$= \underline{0.0439}$$

$$P(B|x=1, Y=1) = \frac{P(x=1|B)\,P(Y=1|B)\,P(B)}{P(x=1, Y=1|A).P(A) + P(x=1, Y=1|B).P(B)} = \frac{0.8\times0.7\times0.7}{0.41}$$

$$= \underline{0.956}$$

since. $P(B|x=1, Y=1)$ is higher, the new instance has higher probability of belonging to class B.

6.

Text.	class	after removing stop words
It was a bad movie	−	bad movie
The movie had no plot	−	movie no plot
It was a great movie	+	great movie
The movie had a good plot	+	movie good plot
the movie was a boring movie.	−	movie boring movie

Text to classify: "The movie had a boring plot"

= movie boring plot = ?

to find

$P(+|\text{movie boring plot})$

$P(-|\text{movie boring plot})$

$P(+ \mid \text{movie boring plot})$

$= P(\text{movie} \mid +) \times P(\text{boring} \mid +) \times P(\text{plot} \mid +) \times P(+)$

$= \frac{2}{5} \times 0 \times \frac{1}{5} \times \frac{2}{5}$

$= 0$

Applying laplace smoothing $5 \to 5+7$

$= \frac{3}{12} \times \frac{1}{12} \times \frac{2}{12} \times \frac{2}{5}$

$= \frac{1}{720}$

word count		
word	+	-
bad	0	1
movie	2	4
no	0	1
plot	1	1
great	1	0
good	1	0
boring	0	1
	5	8

$P(- \mid \text{movie boring plot})$

$= P(\text{movie} \mid -) \times P(\text{boring} \mid -) \times P(\text{plot} \mid -) \times P(-)$

$= \frac{4}{8} \times \frac{1}{8} \times \frac{1}{8} \times \frac{3}{8}$

$= \frac{3}{640} > \frac{1}{720}$.

The test "The movie had a boring plot" will be classified into the "$-$" class.