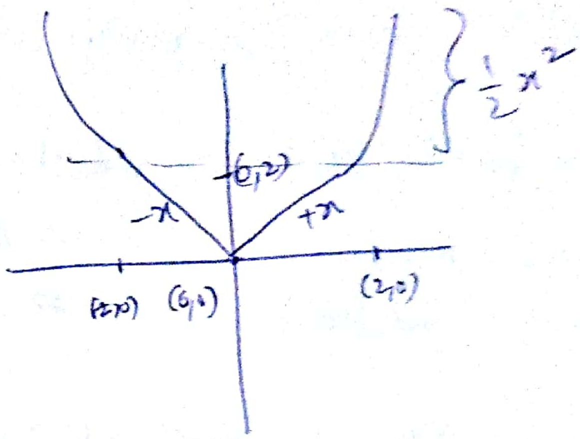
Warm-Up

1) (a) $f(x) = \max\left\{\frac{1}{2}x^2, |x|\right\}$



at point $(0,0)$ $f(x)$ follows $|x|$

$\Rightarrow \quad \partial f(x) = \frac{x}{|x|}$

$= -1 \quad x < 0$

$= 1 \quad x > 0$

So, the set $[-1, 1]$ is the subgradient of $f(x)$ at $(0,0)$

(b) at point $x = -2$

$f(x) = \frac{1}{2}x^2$ from $x \leq -2$ and $x > 2$

$f(x) = |x|$ for $x \geq -2$ and $x < 2$

$\partial(f(x)) = \frac{2x}{2} \quad \forall \; x \leq -2$ and $x > 2$

$= \frac{x}{|x|} \quad \forall \; x > -2$ and $x < 2$

$\Rightarrow$ the set $[-2, -1]$ is the set of subgradients.

2) given $f(x) = \max\{ax+b, cx+d\}$

$f(\lambda x + (1-\lambda)y) = \max\{a(\lambda x + (1-\lambda)y) + b, \; c(\lambda x + (1-\lambda)y) + d\}$

$= \max\{\lambda a x + (1-\lambda)ay + b(\lambda + 1 - \lambda),$
$\lambda c x + (1-\lambda)cy + d(\lambda + 1 - \lambda)\}$

$$= \max \{ \lambda(ax+b) + (1-\lambda)(ay+b) \,,\, \lambda(cx+d) + (1-\lambda)(cy+d) \}$$

$f(x) = \max\{ax+b, cx+d\}$. $\quad f(y) = \max\{ay+b, cy+d\}$

$\Rightarrow \quad ax+b \leq \max(f(x)) \qquad \Rightarrow ay+b \leq \max(f(y))$

$\quad cx+d \leq \max(f(x)). \qquad \Rightarrow cy+d \leq \max(f(y))$

$$\Rightarrow f(\lambda x + (1-\lambda)y) \leq \max\{ \lambda f(x) + (1-\lambda) f(y) \,,\, \lambda f(x) + (1-\lambda) f(y) \}.$$

$$\leq \lambda f(x) + (1-\lambda) f(y).$$

hence

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

Problem 1 :

① 1st Iteration :

$w = [6.9358, 215.9910, 18.6899, 225.5469]$

$b = 212$

2nd Iteration :

$w = [261.1638, 80.8717, -222.8135, 90.9115]$

$b = -182$

3rd Iteration :

$w = [-13.8715, 431.9820, 37.3798, 451.0937]$

$b = 424$

Final weights & bias :

$w = [-753.8585, 770.4289, -535.3486, 817.2905]$

$b = 2$

no. of iterations = 60

② 1st Iteration :

$w = [-0.9474, -0.5776, -0.6663, -0.1427]$

$b = -1$

2nd Iteration :

$w = [-0.1363, 0.3546, 0.0599, 0.5115]$

$b = 0$

3^{rd} Iteration :

$w = [-0.8468,\ 0.2477,\ -0.\cancel{27}41,\ 0.2932]$

$b = -1$

Final Iteration :

$w = [-28.7376,\ 29.2021,\ -20.1797,\ 31.3513]$

$b = 0$

no. of iterations : 114851

Scanned by CamScanner

Problem 1 :

~~(a)~~

(3) Rate of convergence remains constant at 60 in the program. This is attributed due to the fact that w. and b ~~are~~ are changed at the rate of the step size.

For example.

$\eta_t = 1 \quad w_1 = \begin{bmatrix} -753 \\ 770 \\ -535 \\ 817 \end{bmatrix} \quad b_1 = 2$

$\eta_t = 2 \quad w_2 = \begin{bmatrix} -1.5077e^{3} \\ 1.5409e^{3} \\ -1.0707e^{03} \\ 1.6346e^{03} \end{bmatrix} \quad b_2 = 4 \qquad w_2 = 2 \cdot w_1 \quad b_2 = 2 \cdot b_1$

$\eta_t = 1000 \quad w_3 = \begin{bmatrix} -7.5386e^{05} \\ +7.7043e^{05} \\ -5.3535e^{+05} \\ 8.1729e^{+05} \end{bmatrix} \quad b_3 = 2000 \qquad w_3 = 1000 w_1 \quad b_3 = 1000 b_1$

$\eta_t = 0.1 \quad w_4 = \begin{bmatrix} -75.3 \\ 77.04 \\ -53.5 \\ 81.7 \end{bmatrix} \quad b_4 = 0.2 \qquad w_4 = 0.1 w_1 \quad b_4 = 0.1 b_1$

(4)

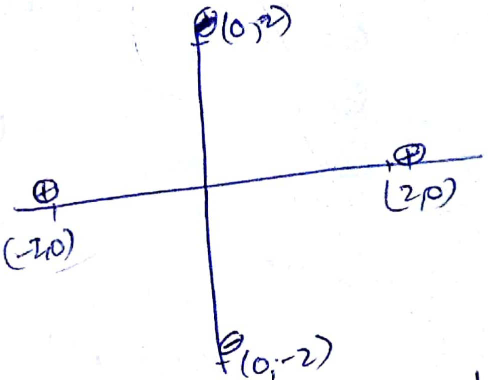
Problem 2:

(a) $\phi(x_1, x_2) = \begin{bmatrix} x_1 + x_2 \\ x_1 - x_2 \end{bmatrix}$.

x_1	x_2	y
-1	-1	1 ⊕
-1	1	-1 ⊖
1	-1	-1 ⊖
1	1	1 ⊕

transformed values:

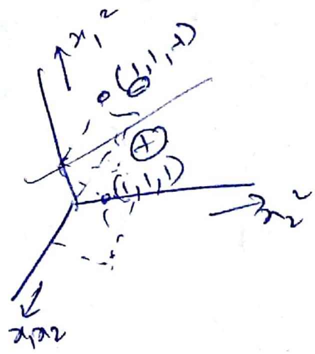
x_1+x_2	x_1-x_2	y
-2	0	⊕
0	-2	⊖
0	2	⊖
2	0	⊕



above points cannot be linearly separated using any 2D transformed vector.

(b) $\phi(x_1, x_2) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$

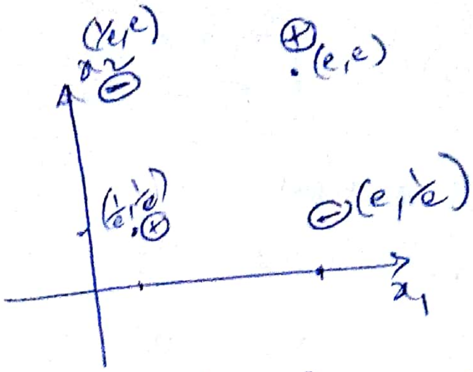
x_1^2	x_2^2	x_1x_2	y
1	1	1	⊕
1	1	-1	⊖
1	1	-1	⊖
1	1	1	+



above points can be linearly separated in the transformed vector space.

(c) $\phi(x_1, x_2) = \begin{bmatrix} e^{x_1} \\ e^{x_2} \end{bmatrix}$

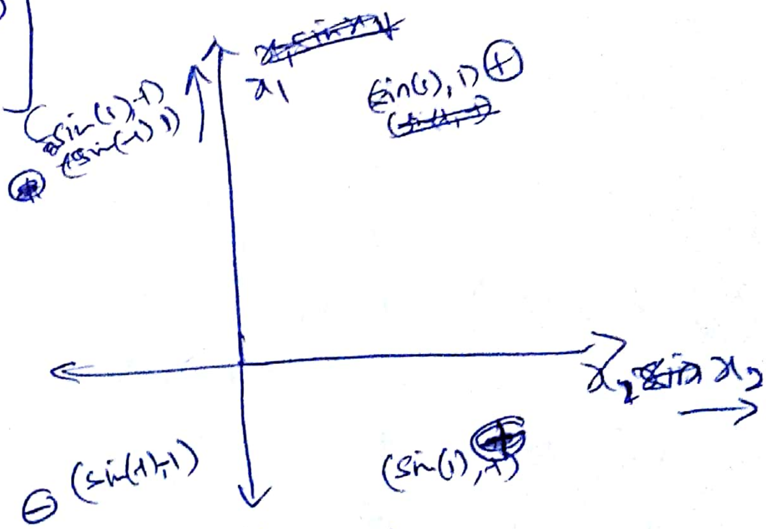
e^{x_1}	e^{x_2}	y
$\frac{1}{e}$	$\frac{1}{e}$	+
$1/e$	e	−
e	$1/e$	−
e	e	+



Using $\phi(x_1, x_2)$, we can linearly separate the data in the above graph plot.

(d) $\phi(x_1, x_2) = \begin{bmatrix} x_1 \sin(x_2) \\ x_1 \end{bmatrix}$

$x_1 \sin(x_2)$	x_1	y
$-1 \sin(-1)$	-1	+
$-1 \sin(1)$	-1	−
$+1 \sin(-1)$	1	−
$1 \sin(1)$	1	+

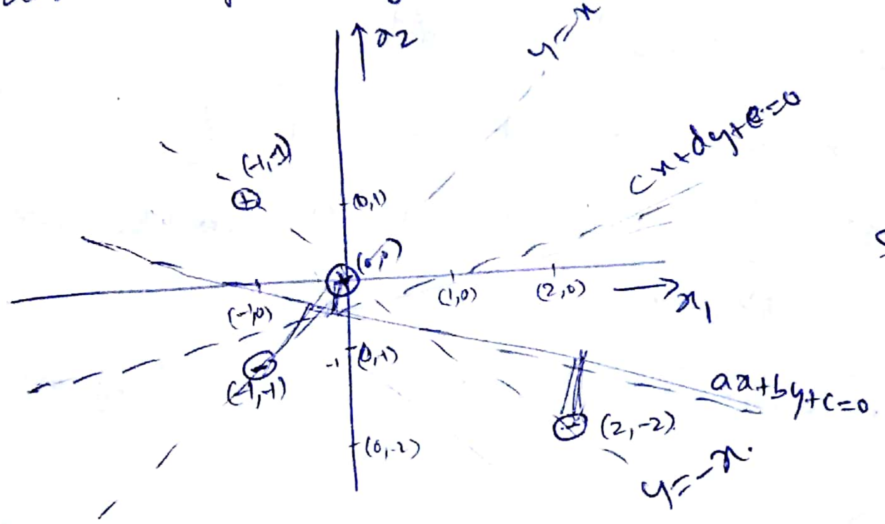


Using $\phi(x_1, x_2)$ data can be linearly separated in the above graph plot

Scanned by CamScanner

Problem 3 :

① Taking 2D plot of the points given ; with ⊕ representing 1 and ⊖ representing -1 :



(-1,1), (0,0), (2,-2) lie on the same line $y=x$.

Support vectors:

Set of data points which satisfy $y^i(w^T x^i + b) = 1$. → ①

the line $y=-x$ is a linear classifier but not perfect — ⓐ

the line $y=x$ is also a linear classifier but not perfect — ⓑ

any line that lies between '$y=x$' and '$y=-x$' can be one of the perfect linear classifiers. and can be represented by $ax+by+c=0$.

⇒ the point (0,0) is at unit distance if $c=1$

⇒ $ax+by+1=0$.

the point closest to (0,0) is (-1,-1), so (-1,-1) can also be considered support vector.

Similarly (2,-2) can also be considered support vector if the slope of line is negative.

for (-1,-1) to be at unit distance :

$(-1)(a(-1)+b(-1)+1)=1$ (from ①)

$\Rightarrow a+b-1=1 \Rightarrow a+b=2$ — ⓒ

for (2,-2) to be at unit distance from perfect classifier.

$(-1)\left(2(a)+(-2)b+1\right)=0$

$\Rightarrow \quad 2b-2a-1=0$

$\Rightarrow \quad b-a=\frac{1}{2}$ —— ⓓ

Solving ⓒ and ⓓ we get $a=\frac{1}{2}$ $b=\frac{3}{2}$

$\Rightarrow \quad \frac{1}{2}(x_1)+\frac{3}{2}(x_2)+1=0.$ can be used as perfect linear classifier

So, $\underline{(0,0),\ (-1,-1)\ \text{and}\ (2,-2)}$ are support vectors here

ⓑ ② let $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be the co-efficients of featurevector $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

Objective : minimize $||w||^2$ such that $y^i(w^T x^i+b)\geq 1$

$\Rightarrow \quad (w_1^2+w_2^2)_{min}$ such that $y^i(w_1x_1^i+w_2x_2^i+b)\geq 1$

Substituting given points we get.

minimize $(w_1^2+w_2^2)$ such that

$b\geq 1 \longrightarrow$ point $(0,0)$

$w_1+w_2-b\geq 1 \longrightarrow$ point $(-1,-1)$

$2w_2-2w_1-b\geq 1 \longrightarrow$ point $(2,-2)$

$w_2-w_1+b\geq 1 \longrightarrow$ point $(-1,1)$

Using lagrangian here

$$L(w_1,w_2,b,\lambda_1,\lambda_2,\lambda_3,\lambda_4)= w_1^2+w_2^2 + \lambda_1(1-b)+\lambda_2(1-(w_1+w_2-b)) + \lambda_3(1-(2w_2-2w_1-b)) + \lambda_4(1-(w_2-w_1+b))$$

Taking derivatives with w_1, w_2 and b:

$$\frac{dL}{dw_1} = 2w_1 + \lambda_2[-1] + \lambda_3[+2] + \lambda_4(+1)$$

$$= 2w_1 - \lambda_2 + 2\lambda_3 + \lambda_4$$

setting it to '0'.

$$\frac{dL}{dw_1} = 0 \Rightarrow 2w_1 = \lambda_2 - 2\lambda_3 - \lambda_4$$

$$w_1 = \frac{\lambda_2 - 2\lambda_3 - \lambda_4}{2} \quad \text{———— (a)}$$

$$\frac{dL}{dw_2} = 2w_2 + \lambda_2[-1] + \lambda_3(-2] + \lambda_4[-1] = 0$$

$$2w_2 = \lambda_2 + 2\lambda_3 + \lambda_4$$

$$w_2 = \frac{\lambda_2 + 2\lambda_3 + \lambda_4}{2} \quad \text{———— (b)}$$

$$\frac{dL}{db} = -\lambda_1 + \lambda_2 + \lambda_3 - \lambda_4 = 0$$

$$\Rightarrow \lambda_2 + \lambda_3 = \lambda_1 + \lambda_4$$

Using dual (g) of L, we get

$$g_{w_1 w_2 b}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \left(\frac{\lambda_2 - 2\lambda_3 - \lambda_4}{2}\right)^2 + \left(\frac{\lambda_2 + 2\lambda_3 + \lambda_4}{2}\right)^2 +$$

$$\lambda_1(1-b) + \lambda_2\left(1 - \left(\frac{\cancel{\lambda_2 - 2\lambda_3 - \lambda_4}}{2} + \frac{\lambda_2 + \cancel{2\lambda_3 + \lambda_4}}{2} - b\right)\right) +$$

$$\lambda_3\left(1 - \left(\cancel{\lambda_2} - 2\lambda_3 - \lambda_4\right) + \left(\cancel{\lambda_2} + 2\lambda_3 + \lambda_4\right) - b\right) +$$

$$\lambda_4\left(1 - \left(\frac{\cancel{\lambda_2} + 2\lambda_3 + \lambda_4}{2} - \frac{\cancel{\lambda_2} - 2\lambda_3 - \lambda_4}{2} + b\right)\right)$$

$$= \frac{(\lambda_2 - 2\lambda_3)^2 + \lambda_4^2 - 2\lambda_4(\lambda_2 - 2\lambda_3)}{4} + \frac{(\lambda_2 + 2\lambda_3)^2 + \lambda_4^2 + 2\lambda_4(\lambda_2 + 2\lambda_3)}{4} +$$

$$\lambda_1(1-b) + \lambda_2(1 - \lambda_2 + b) + \lambda_3(1 - 4\lambda_3 - 2\lambda_4 + b) +$$

$$\lambda_4(1 - (2\lambda_3 + \lambda_4 + b))$$

$$= \frac{\lambda_2^2 + 4\lambda_3^2 - 4\lambda_2\lambda_3 + \lambda_4^2 - 2\lambda_2\lambda_4 + 4\lambda_3\lambda_4}{4} + \frac{\lambda_2^2 + 4\lambda_3^2 + 4\lambda_2\lambda_3 + \lambda_4^2 + 2\lambda_4\lambda_2 + 4\lambda_3\lambda_4}{4}$$

$$+ \lambda_1(1-b) + (\lambda_2 - \lambda_2^2 + b\lambda_2) + \lambda_3 - 4\lambda_3^2 - 2\lambda_4\lambda_3 + b\lambda_3 + \lambda_4 - 2\lambda_3\lambda_4 - \lambda_4^2 - b\lambda_4$$

$$= \lambda_2^2\left[\frac{1}{4} + \frac{1}{4} - 1\right] + \lambda_3^2(1+1-4) + \lambda_4^2\left[\frac{1}{4} + \frac{1}{4} - 1\right] + 2\lambda_3\lambda_4 + \lambda_1(1-b)$$

$$+ \lambda_2(1+b) + \lambda_3(1+b) - 4\lambda_3\lambda_4 + \lambda_4(1-b).$$

* setting $\frac{dg}{d\lambda_1}, \frac{dg}{d\lambda_2}, \frac{dg}{d\lambda_3}, \frac{dg}{d\lambda_4} = 0$ to find infimum.

$$\Rightarrow \frac{dg}{d\lambda_1} = (1-b) = 0$$

$$b = 1 \quad \text{—— (c)}$$

$$\frac{dg}{d\lambda_2} = 2\lambda_2\left[\frac{-1}{2}\right] + (1+b) = 0$$

$$\Rightarrow -\lambda_2 + 2 = 0$$

$$\Rightarrow \lambda_2 = \frac{2}{2}. \quad \text{—— (d)}$$

$$\frac{dg}{d\lambda_3} = -2\lambda_3(2) + 2\lambda_4 + (1-b) - 4\lambda_4 = 0$$

$$\Rightarrow -4\lambda_3 - 2\lambda_4 + 0 = 0$$

$$\Rightarrow (-2)(2\lambda_3 + \lambda_4) = 0$$

$$\Rightarrow 2\lambda_3 + \lambda_4 = 0 \quad \text{—— (e)}$$

~~Using (d) and (e) we get~~

$$\cancel{w_1 = \lambda_2 - \frac{(2\lambda_3 + \lambda_4)}{2} = (2/2) - 0 = (2/2)}$$

$$\cancel{w_2 = \lambda_2 + \frac{(2\lambda_3 + \lambda_4)}{2}}$$

$$\frac{dg}{d\lambda_4} = 2\lambda_4\left(\frac{-1}{2}\right) + 2\lambda_3 - 4\lambda_3 + (1-b) = 0$$

$$\Rightarrow \lambda_4 + 2\lambda_3 = 1 \quad \text{—— (f)}$$

Using (d) and (f)

$$w_1 = \frac{\lambda_2 - (2\lambda_3 + \lambda_4)}{2} = \frac{2-1}{2} = 1/2$$

$$w_2 = \frac{\lambda_2 + (2\lambda_3 + \lambda_4)}{2} = \frac{2+1}{2} = 3/2$$

$b = 1$ from (c)

$\Rightarrow$ $\boxed{y^i = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix} x^i + 1}$ is the perfect linear classifier for given training data.

weight $= [1/2 \cdot 3/2]$

bias $= 1$

(3) Margin is given as $\frac{1}{||w||}$

$$= \frac{1}{\sqrt{(\frac{1}{2})^2 + (\frac{3}{2})^2}} = \frac{1}{\sqrt{\frac{1}{4} + \frac{9}{4}}} = \frac{2}{\sqrt{10}} //$$

Problem 4:

objective: minimize $||w||^2$
such that
$y^i(w^T x^i + b) \geq 1 \quad \forall i \in$ training data.

We use quadprog function to solve the quadratic optimize problem in Matlab.

$w = \text{quadprog}(H, f, A, b)$

Where 'w' is the coefficient weights vector.
'H' is the identity matrix in our solution with last ~~[illegible]~~ matrix value set to 0.

$f =$ zero vector

$A = (-1)(y^i x^i)$
$b = -1$
} quadprog accepts A, b as $A \cdot w \le b$.

let $\begin{bmatrix} w^T \\ b \end{bmatrix}$ be $z \Rightarrow z^T = [w \cdot b]$
our solution is

$$z^T = \text{quadprog}\big(H, \text{zeros}, \underbrace{(-1) .* y^i .* x^i, -1}_{\text{transformation explanation}}\big).$$

$y^i(w^T x^i + b) \geq 1$

$\Rightarrow y^i w^T x^i + y^i b \geq 1$

$\Rightarrow y^i \begin{bmatrix} w^T \\ b \end{bmatrix} \begin{bmatrix} x^i \\ 1 \end{bmatrix} \geq 1$

$\Rightarrow (-1)\, y^i \begin{bmatrix} w^T \\ b \end{bmatrix} \begin{bmatrix} x^i \\ 1 \end{bmatrix} \leq -1 \quad \Rightarrow \quad b = -1$

$$A = (-1) .* y^i .* \begin{bmatrix} w^T \\ b \end{bmatrix} \cdot \begin{bmatrix} x^i \\ 1 \end{bmatrix}$$
$$= (-1) .* y^i .* z .* \begin{bmatrix} x^i \\ 1 \end{bmatrix}$$

after solving we get:

$$\boxed{\begin{aligned} w_1 &= -1.3951 \\ w_2 &= -1.9450 \\ w_3 &= 1.7552 \\ w_4 &= 1.6133 \\ b &= -0.0317 \end{aligned}}$$

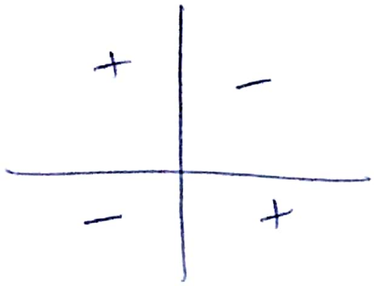
Problem 1:

(4) → With '1' datapoint we can find a linear separator using perceptron.

→ With '2' data points, we can classify

→ With '3' data points, we can classify

→ With 4 data points, certain configurations are not classifiable. For instance



this data set cannot be classified perfectly since any hyperplane doesnt segragate '+' and −

So, minimum no. of points in data set $= 4$

In general:

Whenever given data is not linearly separable using hyperplane, perceptron algorithm fails to converge.

Q Warm Up ①

(b) $g(x) = \max\{\exp(x), 10x\}$

at $x = 2$ $\quad g(x) = \cancel{e} \, 10 \cancel{x} x. = 10x.$

$\Rightarrow \nabla g(x) = 10.$

subgradient $= 10.$

at $x = -1$ $\quad g(x) = \frac{1}{e}$. follows e^{x}.

$\nabla(g(x)) = e^{x}.$

subgradient $= \frac{1}{e}.$