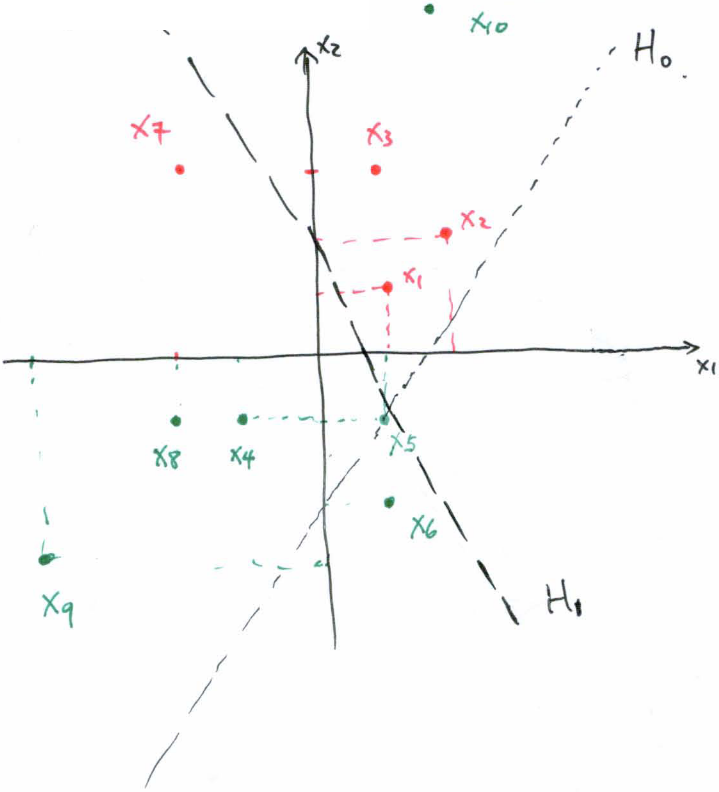
Solution

Q 1.



	x_1	x_2	label
x_1	1	1	w_2
x_2	2	2	w_2
x_3	1	3	w_2
x_4	-1	-1	w_1
x_5	1	-1	w_1
x_6	1	-2	w_1
x_7	-2	3	w_2
x_8	-2	-1	w_1
x_9	-4	-3	w_1
x_{10}	2	6	w_1

(a) $g(x) = w_0 + w_1 x_1 + w_2 x_2$

① $g(x_5) = w_0 + w_1 - w_2 = 0$

② $g(x_2) = w_0 + 2w_1 + 2w_2 = +1$

$$\ominus \begin{cases} 2w_0 + 2w_1 - 2w_2 = 0 \\ w_0 + 2w_1 + 2w_2 = +1 \end{cases}$$

$\Rightarrow \quad w_0 - 4w_2 = -1$

since $w_0 = 1$.

$1 - 4w_2 = -1$

$w_2 = \frac{1}{2}$

from $g(x_5)=0$, $w_0 + w_1 - w_2 = 0$ $\quad w_1 = -\frac{1}{2}$

$g(x) = 1 - \frac{1}{2}x_1 + \frac{1}{2}x_2$

H_0 : $g(x) = 0$

$1 - \frac{1}{2}x_1 + \frac{1}{2}x_2 = 0$

$\frac{1}{2}x_2 = \frac{1}{2}x_1 - 1$

$\underline{x_2 = x_1 - 2}$

$\therefore H_0: \quad x_2 = x_1 - 2.$

(b) :

$g(x_1) = 1 > 0$ $\quad w_1$ $\quad$ should be w_2

$g(x_2) = 1 > 0$ $\quad w_1$ $\quad$ ... w_2

$g(x_3) = 2 > 0$ $\quad w_1$ $\quad$ ... w_2

$g(x_4) = 1 > 0$ $\quad w_1$ $\quad$ correct

$g(x_5) = 0$ $\quad$ Boundary. (assumption).

$g(x_6) = -1/2 < 0$ $\quad w_2$ $\quad$ should be w_1

$g(x_7) = \frac{7}{2} > 0$ $\quad w_1$ $\quad$ should be w_2

$g(x_8) = \frac{3}{2} > 0$ $\quad w_1$ $\quad$ correct.

$g(x_9) = \frac{3}{2} > 0$ $\quad w_1$ $\quad$ correct

$g(x_{10}) = 3 > 0$ $\quad w_1$ $\quad$ correct.

$\therefore$ x_4, x_8, x_9, x_{10} are correct, $\frac{4}{10} = 40\%$.

Q 1 ⓒ

A training set is made from

$X_T = [X_1, X_3, X_4, X_8]$ $\quad\quad \eta(\cdot) = 0.1$ learning rate.

Perceptron : $J_p = \sum_{y \in Y} (-a^T y)$

$$
\begin{array}{cccccc}
 & \underline{w_2} & & \underline{w_1} & \\
 & x_1 & x_3 & x_4 & x_8 \\
y = & \multicolumn{4}{c}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -2 \\ 1 & 3 & -1 & -1 \end{bmatrix}}
\end{array}
\xrightarrow{\text{Normalization}}
\hat{y} = \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -2 \\ -1 & -3 & -1 & -1 \end{bmatrix}
$$

$a(0)$ is from $H(0)$: $g(x) = 1 - \frac{1}{2}x_1 + \frac{1}{2}x_2 \Leftarrow a(0) = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$

$a(0)^T = [1 \;\; -0.5 \;\; 0.5]$

<u>1st Iteration</u> :

$$a(0)^T y = [1 \;\; -0.5 \;\; 0.5] \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -2 \\ -1 & -3 & -1 & -1 \end{bmatrix}$$

$$= [-1 \;\; -2 \;\; 1 \;\; 1.5]$$

misclassified

updating Rule : $a(k+1) = a(k) + \eta(k) \sum_{y \in Y} (y)$ $\quad$ Y: misclassified.

$$a(1) = \begin{pmatrix} 1 \\ -0.5 \\ 0.5 \end{pmatrix} + 0.1 \left[\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ -3 \end{pmatrix} \right]$$

$$= \begin{pmatrix} 1 \\ -0.5 \\ 0.5 \end{pmatrix} + 0.1 \begin{pmatrix} -2 \\ -2 \\ -4 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8 \\ -0.7 \\ 0.1 \end{pmatrix}$$

iter 2

$$a(1)^T y = [0.8 \ \ -0.7 \ \ 0.1] \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -2 \\ -1 & -3 & -1 & -1 \end{bmatrix}$$

$$= [-0.2 \ \ -0.4 \ \ 1.4 \ \ 2.1]$$

misclassified.

$$a(2) = \begin{pmatrix} 0.8 \\ -0.7 \\ 0.1 \end{pmatrix} + 0.1 \left[\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ -3 \end{pmatrix} \right]$$

$$= \begin{pmatrix} 0.8 \\ -0.7 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 \\ -0.2 \\ -0.4 \end{pmatrix}$$

$$= \begin{pmatrix} 0.6 \\ -0.9 \\ -0.3 \end{pmatrix}$$

iter 3

$$a(2)^T y = [0.6 \ \ -0.9 \ \ -0.3] \begin{bmatrix} -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -2 \\ -1 & -3 & -1 & -1 \end{bmatrix}$$

$$= [0.6 \ \ 1.2 \ \ 1.8 \ \ 2.7].$$

All classified well.

$$\text{Final } a = \begin{pmatrix} 0.6 \\ -0.9 \\ -0.3 \end{pmatrix}$$

$$H_1 : \quad 0.6 - 0.9x_1 - 0.3x_2 = 0$$

$$x_2 = -3x_1 + 2$$

Q1 (d) single-sample method after seeing all 4 training samples just once, will yield result as good as iteration 1 of the batch method in part (c).

Q 2.

It seems that LMS is more suitable, b/c.

(1) given 10 samples are not linearly separable.

(2) Relaxation method in (i) will always be in error, which will never settle.

Q③

(a)

x_1	1	1	w_2
x_2	2	2	w_2
x_4	-1	-1	w_1
x_7	-2	3	w_2
x_8	-2	-1	w_1

w_2 w_1

$$
\begin{array}{cccccc}
 & x_1 & x_2 & x_4 & x_7 & x_8 \\
y = & \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & -1 & -2 & -2 \\ 1 & 2 & -1 & 3 & -1 \end{bmatrix}
\end{array}
$$

$$(\hat{y})^T = \begin{bmatrix} -1 & -1 & 1 & -1 & 1 \\ -1 & -2 & -1 & 2 & -2 \\ -1 & -2 & -1 & -3 & -1 \end{bmatrix}$$

$$\hat{y} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -2 & -2 \\ 1 & -1 & -1 \\ -1 & 2 & -3 \\ 1 & -2 & -1 \end{bmatrix}$$

$$(\hat{y})^T \hat{y} = \begin{pmatrix} -1 & -1 & 1 & -1 & 1 \\ -1 & -2 & -1 & 2 & -2 \\ -1 & -2 & -1 & -3 & -1 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ -1 & -2 & -2 \\ 1 & -1 & -1 \\ -1 & 2 & -3 \\ 1 & -2 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 5 & -2 & 4 \\ -2 & 14 & 2 \\ 4 & 2 & 16 \end{pmatrix}$$

$$|(\hat{y})^T\hat{y}| = \begin{pmatrix} 5 & -2 & 4 \\ -2 & 14 & 2 \\ 4 & 2 & 16 \end{pmatrix} = 5\cdot 14\cdot 16 + (-2)\cdot 2\cdot 4 + 4\cdot(-2)\cdot(2) - 4\cdot 14\cdot 4 - 2\cdot 2\cdot 5 - 16\cdot(-2)(-2)$$

$$= 1120 - 16 - 16 - 224 - 20 - 64$$

$$= 780.$$

$$\left((\hat{y})^T \hat{y}\right)^{-1} = \frac{1}{780}\begin{pmatrix} 220 & 40 & -60 \\ 40 & 64 & -18 \\ -60 & -18 & 66 \end{pmatrix}$$

$$a = \left((\hat{y})^T \hat{y}\right)^{-1} \hat{y}^T b$$

since $b = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$

$$a = \frac{1}{780}\begin{pmatrix} 100 \\ -152 \\ -396 \end{pmatrix} \Rightarrow g(x) = \frac{100}{780} - \frac{152}{780}x_1 - \frac{396}{780}x_2$$

$$\frac{100}{780} - \frac{152}{780}x_1 - \frac{396}{780}x_2 = 0$$

$$0.1282 - 0.1949x_1 - 0.5077x_2 = 0$$

$$x_2 = -0.3838x_1 + 0.2525 \Longrightarrow H_3$$

Q3 (b) $g(x) = 0.1282 - 0.1949x_1 - 0.5077x_2.$

$$g(x_5) = 0.1282 - 0.1949(1) - 0.5077(-1)$$
$= 0.4410$ $\quad x_5 \in w_1$

$$g(x_7) = 0.1282 - 0.1949(-2) - 0.5077(3)$$
$= -1.0051$ $\quad x_7 \in w_2$

$$g(x_9) = 0.1282 - 0.1949(-4) - 0.5077(-3)$$
$= 2.4309$ $\quad x_9 \in w_1$

$g(X_{10}) = 0.1282 - 0.1949(2) - 0.5077(6)$

$= -3.3078 \qquad X_{10} \in w_2$ in error

All samples classified correctly, except X_{10} is in error.

$\therefore H_3$ achieves a good compromise, such that it translates into a better generalization on unseen data in terms of classification error.