Name: Anurag Shinde

HOMEWORK 1

Exercise 1.3 The weight update rule in {1.3) has the nice interpretation that it moves in the direction of classifying x(t) correctly.

(a) Show that y(t)wT(t)x(t) < 0. [Hint: x(t) is misclassified by w(t).]

Response: Since x(t) is misclassified by w(t) hence y(t) = 1 for sign(wTx) = -1 which implies wTx < 0 hence y(t)wT(t)x(t) < 0

(b) Show that y(t)wT(t+l)x(t) > y(t)wT(t)x(t). [Hint: Use (1.3).]

Response: Since w(t+1) = w(t) + y(t)x(t) where t the iteration number.
This proves that w(t+1) > w(t)
Hence product of y(t)w(t)x(t) > y(t)w(t)x(t)

(c) As far as classifying x(t) is concerned, argue that the move from w(t) to w(t+ 1) is a move 'in the right direction'.

Response: Let us suppose x(t) is misclassified. So there are two possible cases:

1. Either x(t) is misclassified positively
2. Or x(t) is misclassified negatively

For mistakes on positive w(t+1) = w(t) + x(t)y(t)

For mistakes on negative w(t+1) = w(t) – x(t)y(t)

So, in both cases we are reduced the mistakes using above equations and move closer to the value that we want.

Exercise 1.6

For each of the following tasks, identify which type of learning is involved (supervised, reinforcement, or unsupervised) and the training data to be used. If a task can fit more than one type, explain how and describe the training data for each type.

(a) Recommending a book to a user in an online bookstore

Response:

- This is a case of supervised learning and the dataset for this problem will be a dataset of previous records where other users have bought similar books on some topics. This is a supervised learning example as we already know some pattern

about what books are bought by previous buyers and hence we can predict recommended books to the new user.
The training dataset may contain following parameter

(a) Book Author
(b) Book Context
(c) Book Size
(d) Book price
(e) Book ratings
(f) Book purchase history
(g) Output parameter will be the status of book sold or not sold.

(b) **Playing tic tac toe**

Response: This is a case of reinforcement learning as we are not sure about the exact output, but we have some previous knowledge where we know how placing marks at a particular position can result in victory or defeat. The data parameters for this will be

- Distance between two consecutive marks
- Probability that two consecutive marks will form a diagonal when merged with a third mark
- Probability that two consecutive marks will form a vertical straight line when merged with a third mark
- Probability that two consecutive marks will form a horizontal straight line when merged with a third mark.

(c) Categorizing movies into different types

Response: This is a case of supervised learning if we have labelled data for categorizing the movies. For example, data parameter will be

- ♦ Movie content (Content of movie determines whether it is for adults or kids but we have an output parameter for content category.)
- ♦ Movie actors (Actors can be used to categorize movie by actions)
- ♦ Movie theme
- ♦ Movie length (Short or long movie but we have an output parameter for classifying movie by length.)
- ♦ Movie title (Know the meaning, we have a model for classifying movie based on title)
- ♦ Date of release (To categorize whether it is old or new movie, we have a threshold value for new and old movie dates)
- ♦ Movie language (To categorize if it is a regional or global movie, we have classification based on language)

(d) Learning to play music

Response: This is a case of reinforcement learning. For example, if we are learning to play music using Guitar then we don't know much about the rhythm of playing the notes but we have some understanding of which notes sound higher and which notes sound lower. It is through practice and observance we will learn how to select best notes that enhances the rhythms and does not breaks it. Data parameters for this learning will be

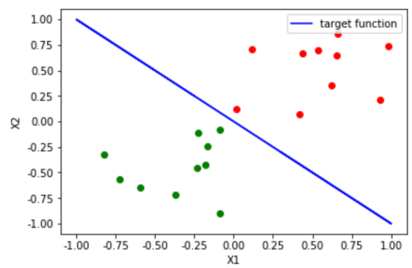
- Frequency of notes
- Tone of each note
- Order of notes played
- Rank of notes

(e) **Credit limit**: Deciding the maximum allowed debt for each bank customer
Response: This is a case of supervised learning as we will have some previous data from other customers based on their salary, bank balance, current loans, number of dependents which can be used to set a max threshold for each individual bank customer. The following are the data parameters that can be used:

a. Age of the customer
b. Customer salary
c. Customer current loans
d. Customer credit score
e. Number of dependents on the customer.

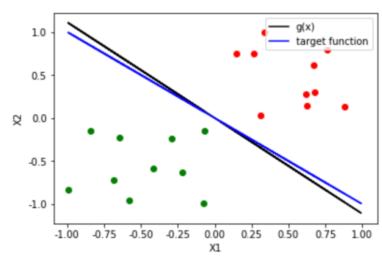
Problem 1.4

a)



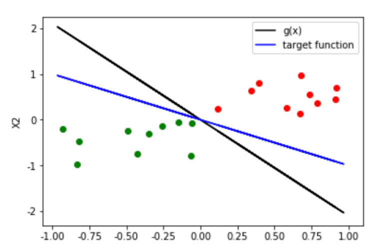
Target function for 20 randomly generated datasets.

b)



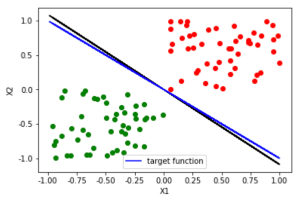
Target function f vs g(x) final hypothesis function. With 20 data sets the two functions are quite close to each other and intersect each other.

c)



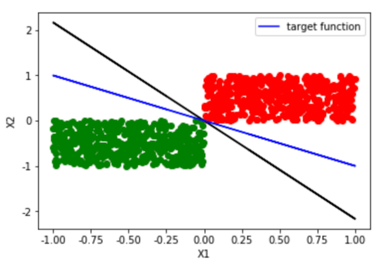
Target function f vs g(x) final hypothesis function. With another 20 data sets the two functions are deviated more as compared to the graph obtained in 1.4 b and intersect each other. This is probably because of the random nature of data.

d)



Target function f vs g(x) final hypothesis function. With another 100 data sets the two functions are quite close to each other with little deviation and intersect each other.

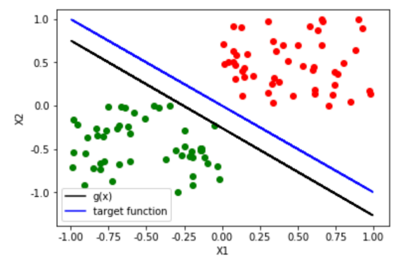
e)



Target function f vs g(x) final hypothesis function. With another 1000 data sets the two functions are deviated by a big gap between the two functions and intersect each other.

Problem 1.5

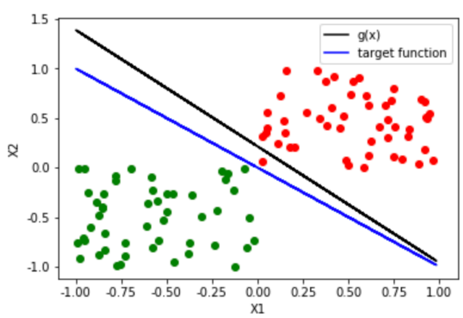
a)



n = 100

Error rate : 1.05%

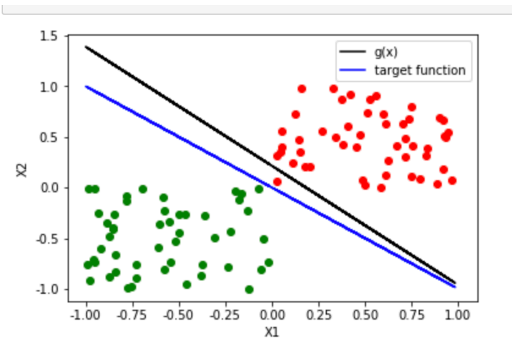
1.5 b)



n = 1

Error rate: 1.05%

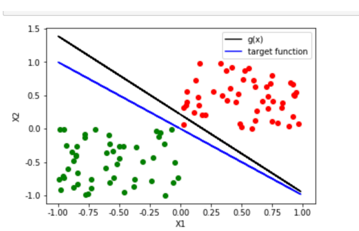
1.5 c)



n = 0.01

Error rate : 1.05%

1.5 d)



n = 0.01

Error rate : 1.05%

1.5 e)

In plot a) the target function f and final hypothesis g(x) separate data linearly and run parallel to each other. In plots b – e the two functions appear to interest in positive direction while there are some data points are misclassified.

Problem 1.2

$$h(x) = \text{sign}(\vec{w}^T \vec{x})$$

$$= \text{sgn}(g(x))$$

Here $x_2 = ax_1 + b$

a = slope of line

when $a=0$ then the above line is horizontal as slope $=0$.

For $a > 0$ it is increasing & b is the offset (y-intercept). Hence we intersect x_2 axis at $x_2 = b$ & line will pass through origin if and only if $b=0$.

$\therefore \quad x_2 = ax_1 + b.$

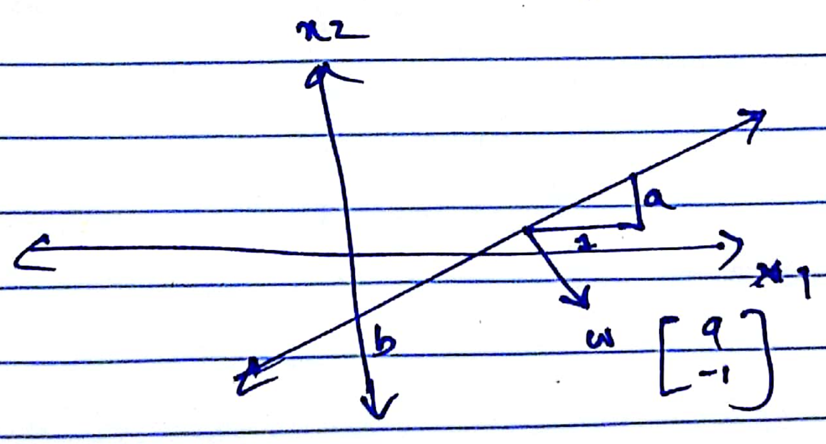
$\Rightarrow \quad -ax_1 + x_2 + b = 0$

converting in vector form

$$[a, -1] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b = 0$$

Considering $x_0 = 1$

$$[b \;\; -a \;\; 1] \cdot \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0$$



linear function is defined by.

$$y(x) = 0 \quad \& \quad y(x) = w^T x$$

$$\therefore (w_0, w_1, w_2) \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0.$$

$$\therefore w_0 + w_1 x_1 + w_2 x_2 = 0.$$

This is the hyperplane with normal vector w

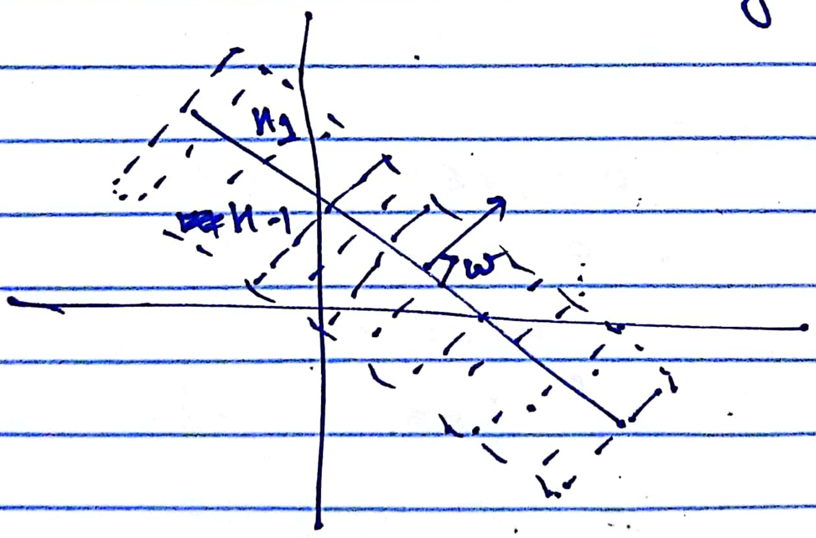
This hyperplane seperates space into two halfs $h(x) = 1$ & $h(x) = -1$

i.e

$$H_{+1} = \{x \mid y(x) > 0\} \quad \& \quad H_{-1} \{x \mid y(x) < 0\}$$

Hence $h(x) = sgn(y(x)) = 1$ for $x \in H_1$

& $h(x) = sgn(yn) = -1$ for $x \in H_{-1}$



b) Draw the picture for cases $\vec{w} = [1, 2, 3]^T$ &
$\vec{w} = -[1, 2, 3]^T$

Case 1:

$$y(x) = 0.$$
$$\Rightarrow w^T x = 0.$$
$$(1, 2, 3) \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0.$$

$(2, 0)$

$(0, 3)$ $= \frac{y_2 - y_1}{x_2 - x_1}$

$= \frac{3}{2}$ $m = y/x$

$y = -1/3$

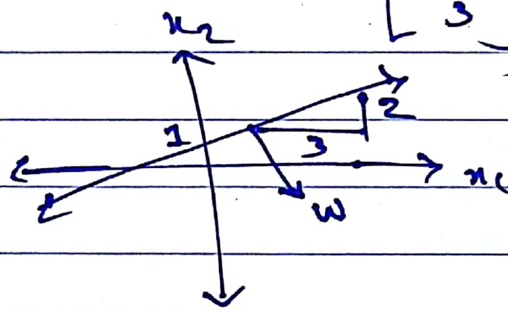
$x = -1/2$

$m = y/x =$

$$1 + 2x_1 + 3x_2 = 0$$
$$2x_1 + 3x_2 = -1$$
$$\therefore w = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

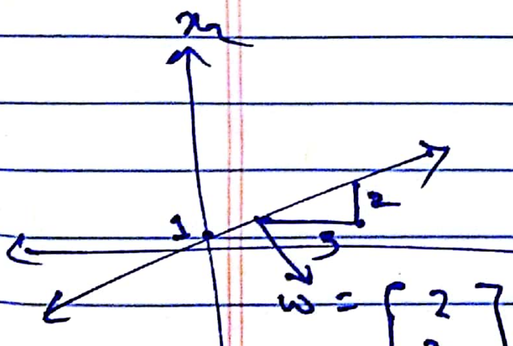
$x_1 = -1/2$ $x_2 = 0$

$x_2 = -1/3$ $x_1 = 0$



Case 2:

$$y(x) = 0$$
$$\therefore w^T(x) = 0.$$
$$[-1, -2, -3] \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix} = 0$$
$$-1 - 2x_1 - 3x_2 = 0$$
$$2x_1 + 3x_2 + 1 = 0$$
$$2x_1 + 3x_2 = -1$$
$$\therefore [2, 3] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + 1 = 0$$
$$\therefore w = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$



$w = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

Scanned by CamScanner

Exercise 1.8

ν = Fraction of red balls

$\therefore P(\nu \le 0.1) = P\left(\frac{red}{N} \le 0.1\right)$

$\Rightarrow P(red \le 0.1N)$

$N=10$

$P(red \le 0.1N) = P(red \le 1) = P(red=0) + P(red=1)$ → ①

$\mu = 0.9 = P(red)$ $\therefore P(green) = 1 - 0.9 = 0.1$

$P(red=0) = \cancel{10 \times 0.1 \times 10^{-10}}\ 10 \times (0.1)^{10} = 10^{-10}$

$P(red=1) = 10 \times 0.1^{9} \times 0.9 = 9 \times 10^{-9}$

$\therefore$ From Eqn ①

$P(red \le 1) = 10^{-10} + 9 \times 10^{-9}$

$= 0.1 \times 10^{-9} + 9 \times 10^{-9}$

$= 9.1 \times 10^{-9}$

$\therefore P(\nu \le 0.1) = 9.1 \times 10^{-9}$

Problem 1.11

For supermarket we have to penalize more if a true customer is reject. (we are not considering correct case as error = 0 in this case).
General form of Ein is given by

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^{N} e(h(x_n), f(x_n)).$$

Penalizing more ⇒

$$= \frac{1}{N} \left[\sum_{y_n=1} e(h(x_n), 1) + \sum_{y_n=-1} e(h(x_n), -1) \right]$$

$$= \frac{1}{N} \left[\sum_{y_n=1} 10 [h(x) \neq 1] + \sum_{y_n=-1} [h(x) \neq -1] \right]$$

For ~~super market~~ CIA we have to penalize more if algorithm accepts an intruder (we are not considering correct case as error = 0)
∴ General form of Ein is given by

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^{N} e(h(x_n), f(x_n))$$

$$= \frac{1}{N} \left[\sum_{y_n=1} e(h(x_n), 1) + \sum_{y_n=-1} e(h(x_n), -1) \right]$$

$$= \frac{1}{N} \left[\sum_{y_n=1} [h(x_n) \neq 1] + \sum_{y_n=-1} 1000 [h(x) \neq -1] \right]$$

Exercise 1.9

General formula

$$P(|\nu-\mu| > \epsilon) \le 2e^{-2\epsilon^2 N} \rightarrow (1)$$

Here $\mu = 0.9$

$\nu = 0.1$

$N = 10$

$\Rightarrow \cancel{|\nu-\mu| \le -0.8}$ $\quad \nu - \mu \le -0.8$

$\therefore \quad |\nu - \mu| \geq 0.8$

$\therefore$ we get $\epsilon = 0.8$ by comparing with above form (1)

$$\therefore P(|\nu-\mu| > 0.8) \le 2e^{-2\times 0.8^2 \times 10}$$

$$\approx 5.52 \times 10^{-6}$$

From solution of ~~exer~~ exercise 1.8 we get $P(\nu \le 0.1) \le P(|\nu-\mu| > 0.8)$

$$\Rightarrow 9.1 \times 10^{-9} \le 5.52 \times 10^{-6}$$

Hence we prove that