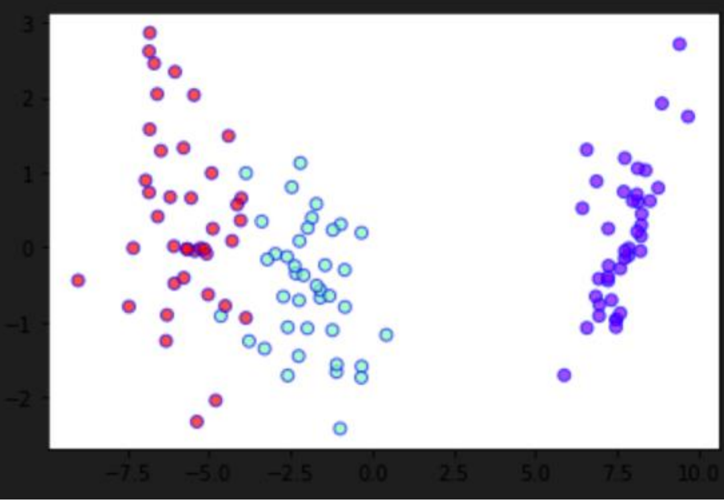
2022S CS 559-B: Homework 1

Name: Janmejay Mohanty

Solution 1:

```
<matplotlib.collections.PathCollection at 0x1a67b90a050>
```




```
Accuracy : 1.0
[[10  0  0]
 [ 0 10  0]
 [ 0  0 10]]
```

Solution 2:

(1) Using Probabilistic Generative Model Equation:

$$P(C_1|x_n) = \sigma(\omega^T x_n) = f(x_n) \qquad \text{------(1)}$$

Using the maximum likelihood to determine the parameters of the logistic regression model,

Here,

$\{x_n, \ y_n\}, n = 1, \ldots. \ , N$

$y_n = 1$ denotes class C_1;

$y_n = 0$ denotes class C_2;

We want to find out the values of w that can maximize the posterior probabilities associated to the observed data.

Therefore, the likelihood function:

$$L(w) = \prod_{n=1}^{N} P(C_1|x_n)^{y_n} * \left(1 - P(C_1|x_n)\right)^{1-y_n}$$

$$= \prod_{n=1}^{N} f(x_n)^{y_n} * \left(1 - f(x_n)\right)^{1-y_n} \qquad \text{--------(2)}$$

Taking the negative logarithm of the likelihood function on equation (2) [Cross Entropy]:

$$\varepsilon(\omega) = -\ln L(\omega) = -\ln \prod_{n=1}^{N} f(x_n)^{y_n} * \left(1 - f(x_n)\right)^{1-y_n}$$

$$= -\sum_{n=1}^{N}\left(y_n \ln f(x_n) + \left(1 - y_n \ln\left(1 - f(x_n)\right)\right)\right) \qquad \text{-----(3)}$$

Therefore:

$$\max L(\omega) = \min \varepsilon(\omega)$$

From equation (1) and (3),

$$\varepsilon(\omega) = -\sum_{n=1}^{N}\left(y_n \ln\left(\sigma(\omega^T x_n)\right) + (1 - y_n)\ln\left(1 - \sigma(\omega^T x_n)\right)\right) \qquad \text{------(4)}$$

Taking derivative of equation (4) with respect to ω,

$$\frac{\partial(\varepsilon(\omega))}{\partial\omega} = \frac{\partial\left[-\sum_{n=1}^{N}\left(y_n \ln\left(\sigma(\omega^T x_n)\right) + (1-y_n)\ln\left(1-\sigma(\omega^T x_n)\right)\right)\right]}{\partial\omega}$$

$$= -\sum_{n=1}^{N}\left[y_n * \left(\frac{1}{\sigma(\omega^T x_n)}\right) * \frac{\partial\left(\sigma(\omega^T x_n)\right)}{\partial\omega} + (1 - y_n) * \left[\frac{1}{1-\sigma(\omega^T x_n)}\right] * \left[-\frac{\partial\left(\sigma(\omega^T x_n)\right)}{\partial\omega}\right]\right] \qquad \text{--(5)}$$

Also,

$$\frac{\partial\left(\sigma(\omega^T x_n)\right)}{\partial\omega} = \frac{\partial\left(\sigma(\omega^T x_n)\right)}{\partial(\omega^T x_n)} * \frac{\partial(\omega^T x_n)}{\partial\omega}$$

$$\frac{\partial\left(\sigma(\omega^T x_n)\right)}{\partial\omega} = \sigma(\omega^T x_n) * \left(1 - \sigma(\omega^T x_n)\right) * x_n \qquad \text{------(6)}$$

From equation (5) and (6),

$$\frac{\partial(\varepsilon(\omega))}{\partial\omega} = -\sum_{n=1}^{N}\left[\frac{y_n}{\sigma(\omega^T x_n)} * \sigma(\omega^T x_n) * \left(1 - \sigma(\omega^T x_n)\right) * x_n - (1 - y_n) * \left[\frac{1}{1-\sigma(\omega^T x_n)}\right] * \sigma(\omega^T x_n) * \left(1 - \sigma(\omega^T x_n)\right) * x_n\right]$$

$$= -\sum_{n=1}^{N}\left[y_n x_n\left(1 - \sigma(\omega^T x_n)\right) - (1 - y_n) * \sigma(\omega^T x_n) * x_n\right]$$

$$= -\sum_{n=1}^{N}\left[y_n x_n - \sigma(\omega^T x_n)(y_n x_n) - \sigma(\omega^T x_n)x_n + \sigma(\omega^T x_n)(y_n x_n)\right]$$

$$= -\sum_{n=1}^{N}\left[y_n x_n - \sigma(\omega^T x_n)x_n\right]$$

$$= \sum_{n=1}^{N}\left[\sigma(\omega^T x_n)x_n - y_n x_n\right]$$

$$= \sum_{n=1}^{N}\left[\sigma(\omega^T x_n) - y_n\right] * x_n \qquad \text{-----------(7)}$$

From equation (1) and (7),

$$\frac{\partial(\varepsilon(\omega))}{\partial\omega} = \sum_{n=1}^{N}\left[f(x_n) - y_n\right] * x_n$$

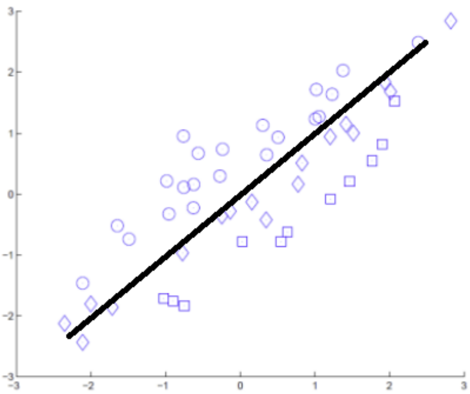
$$=> \quad \boldsymbol{\nabla_\omega \varepsilon(\omega) = \sum_{n=1}^{N}[f(x_n) - y_n] * x_n}$$

(2)

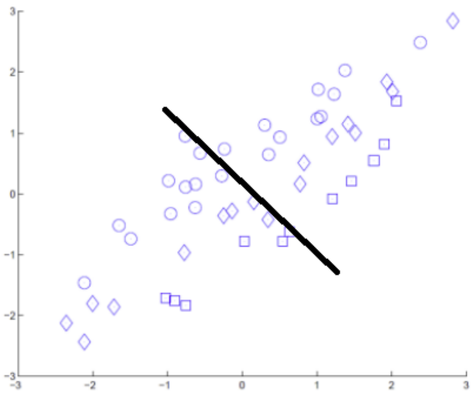
```
Accuracy on test set by our model           :    91.22807017543859
Accuracy on test set by sklearn Logistic Regression model    :    94.73684210526315
```

Solution 3:

(1)



1(a) This one is First PCA component



1(b) This one is Fisher's Linear Discriminant

(2) (a)

X	Y
2	2
0	0
-2	-2

$$\overline{X}(Mean\ of\ X) = \frac{2+0+(-2)}{3}$$
$$= \frac{0}{3}$$
$$\overline{X}(Mean\ of\ X) = 0$$

$$\overline{Y}(Mean\ of\ Y) = \frac{2+0+(-2)}{3}$$
$$= \frac{0}{3}$$
$$\overline{Y}(Mean\ of\ Y) = 0$$

Covariance Matrix

$$C = \begin{bmatrix} Cov(x,y) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix}_{2X2}$$

$$Cov(X,X) = Var(X) = \frac{1}{N-1}\sum_{n=1}^{N}\left(X_n - \overline{X}\right)^2$$
$$Cov(X,Y) = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \overline{X})(Y_n - \overline{Y})$$

X_n	$X_n - \overline{X}$	$\left(X_n - \overline{X}\right)^2$	Y_n	$Y_n - \overline{Y}$	$\left(Y_n - \overline{Y}\right)^2$
2	$2-0$	4	2	$2-0$	4
0	$0-0$	0	0	$0-0$	0

-2	$-2-0$	4	-2	$-2-0$	4

$N = 3$

$Var(X) = \frac{4+0+4}{3-1}$
$= \frac{8}{2}$
$= 4$

$Var(Y) = \frac{4+0+4}{3-1}$
$= \frac{8}{2}$
$= 4$

X_n	Y_n	$X_n - \overline{X}$	$Y_n - \overline{Y}$	$(X_n - \overline{X})(Y_n - \overline{Y})$
2	2	$2-0$	$2-0$	4
0	0	$0-0$	$0-0$	0
-2	-2	$-2-0$	$-2-0$	4

$$Cov(Y,X)\ or\ Cov(X,Y) = \frac{4+0+4}{3-1}$$
$$= \frac{8}{2}$$
$$= 4$$

$$=> \ C = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$
$$C - \lambda \mathrm{I} = 0 \mathrm{Now}$$

Here,
C represents Covariance Matrix.
λ represents Eigen Values.
I represents Identity Matrix.

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 0$$
$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = 0$$
$$\begin{bmatrix} 4-\lambda & 4 \\ 4 & 4-\lambda \end{bmatrix} = 0$$

Taking the determinant of matrix,

$$\begin{vmatrix} 4-\lambda & 4 \\ 4 & 4-\lambda \end{vmatrix} = 0$$
$$(4-\lambda)(4-\lambda) - 4 * 4 = 0$$
$$(4-\lambda)^2 - 16 = 0$$
$$16 + \lambda^2 - 8\lambda - 16 = 0$$
$$\lambda^2 - 8\lambda = 0$$
$$\lambda(\lambda - 8) = 0$$

$\lambda - 8 = 0$ or $\lambda = 0$

$=> \lambda_1 = 8$ $=> \lambda_2 = 0$

Now, we are finding the Eigen Vector's for each Eigen Value's:

$$CW = \lambda W$$

Here,

C represents the Covariance Matrix.

W represents the Eigen Vector.

λ represents the Eigen Values.

For $\lambda_1 = 8$,

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = 8 \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$\begin{bmatrix} 4x_1 & 4y_1 \\ 4x_1 & 4y_1 \end{bmatrix} = \begin{bmatrix} 8x_1 \\ 8y_1 \end{bmatrix}$$

$=> 4x_1 + 4y_1 = 8x_1$

$=> -4x_1 + 4y_1 = 0$ --------------(1)

$=> 4x_1 + 4y_1 = 8y_1$

$=> 4x_1 - 4y_1 = 0$ --------------(2)

From equation (1) and (2), we can say that:

$=> 4x_1 = 4y_1$

$=> \boldsymbol{x_1 = y_1}$ ---------------(3)

Taking $x_1 = 1$, then $y_2 = 1$:

$$=> W_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

$$=> W_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Finding the square root of the sum of squares of the elements in W_1 matrix,

$$\sqrt{(1)^2 + (1)^2} = \sqrt{2} = 1.41$$

Now, dividing the elements of W_1 matrix with 1.41 value:

$$W_1 = \begin{bmatrix} \frac{1}{1.41} \\ \frac{1}{1.41} \end{bmatrix} = \begin{bmatrix} 0.709 \\ 0.709 \end{bmatrix} = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}$$

So, the Eigen Vector are $x_1 = 0.71$ and $y_1 = 0.71$

For $\lambda_2 = 0$,

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = 0 \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$\begin{bmatrix} 4x_2 & 4y_2 \\ 4x_2 & 4y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$=> 4x_2 + 4y_2 = 0$ ---------------------------(4)

$=> 4x_2 + 4y_2 = 0$ ---------------------------(5)

From equation (4) and (5), we can say that:

$=> 4x_2 = -4y_2$

$=> \boldsymbol{x_2 = -y_2}$ ---------------------------(6)

Taking $y_2 = 1$, then $x_2 = -1$:

$$=> W_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$=> W_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Finding the square root of the sum of squares of the elements in W_2 matrix,

$$\sqrt{(-1)^2 + (1)^2} = \sqrt{1+1} = \sqrt{2} = 1.41$$

Now, dividing the elements of W_2 matrix with 1.41 value:

$$W_2 = \begin{bmatrix} \frac{-1}{1.41} \\ \frac{1}{1.41} \end{bmatrix} = \begin{bmatrix} -0.709 \\ 0.709 \end{bmatrix} = \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix}$$

So, the Eigen Vector are $x_2 = -0.71$ and $y_2 = 0.71$.

$W_1 = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}$ $W_2 = \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix}$

$\lambda_1 = 8$ $\lambda_2 = 0$

As, $\boldsymbol{\lambda_1 > \lambda_2}$,

The Eigen Vector associated with the largest Eigen Value corresponds to the first principal component; the Eigen Vector associated with the second largest Eigen Value corresponds to the second principal component.

Therefore, the first principal component is $\boldsymbol{W_1 = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}}$ and $\boldsymbol{\lambda_1 = 8}$.

(b)

X_n	Y_n
2	2
0	0
-2	-2

$$\overline{X} = \frac{2 + 0 + (-2)}{3} = 0$$

$$\overline{Y} = \frac{2 + 0 + (-2)}{3} = 0$$

First Principal Component,

$W_1 = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}$ $\lambda_1 = 8$

	P_{11}	P_{12}	P_{13}
PC_1	2.84	0	−2.84

$$P_{11} = W_1^T \begin{bmatrix} X_n - \overline{X} \\ Y_n - \overline{Y} \end{bmatrix} = [0.71 \quad 0.71] \begin{bmatrix} 2-0 \\ 2-0 \end{bmatrix}$$

$$= [0.71 \quad 0.71] \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$= [1.42 \quad 1.42]$$

$$= 2.84$$

$$P_{12} = W_1^T \begin{bmatrix} X_n - \overline{X} \\ Y_n - \overline{Y} \end{bmatrix} = [0.71 \quad 0.71] \begin{bmatrix} 0-0 \\ 0-0 \end{bmatrix}$$

$$= [0.71 \quad 0.71] \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$= [0 \quad 0]$$

$$= 0$$

$$P_{13} = W_1^T \begin{bmatrix} X_n - \overline{X} \\ Y_n - \overline{Y} \end{bmatrix} = [0.71 \quad 0.71] \begin{bmatrix} -2-0 \\ -2-0 \end{bmatrix}$$

$$= [0.71 \quad 0.71] \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$= [-1.42 \quad -1.42]$$

$$= -2.84$$

	P_{11}	P_{12}	P_{13}
PC_1	2.84	0	−2.84

Mean $(\mu) = \frac{2.84+0+(-2.84)}{3} = 0$

Using Unbiased Variance,

$$Var(PC) = \frac{1}{N-1} \sum_{n=1}^{N} \left(PC_n - \overline{PC}\right)^2$$

$$= \frac{1}{3}[(2.84-0)^2 + (0-0)^2 + (-2.84-0)^2]$$

$$= \frac{1}{2}[(2.84)^2 + (-2.84)^2]$$

$$= \frac{1}{2}[8.0656 + 8.0656]$$

$$= \frac{16.1312}{2}$$

$$= 8.0656$$

(c)

We have calculated eigen value corresponding to the first principal component, that is 8.

So, the cumulative explained variance of the first principal component is % total variance $\left(\frac{8}{8.06} + 0\right) = 0.99.$

As it lies between 90-95%, therefore we can say that variance of the data is captured.

Solution 4:

(1)

The equation of the SVM hyperplain $h(x)$:

$$h(x) = \omega^T x + b$$

i	x_i	y_i	α_i
1	$(4, 2.9)$	1	0.414
2	$(4, 4)$	1	0
3	$(1, 2.5)$	-1	0
4	$(2.5, 1)$	-1	0.018
5	$(4.9, 4.5)$	1	0
6	$(1.9, 1.9)$	-1	0
7	$(3.5, 4)$	1	0.018
8	$(0.5, 1.5)$	-1	0
9	$(2, 2.1)$	-1	0.414
10	$(4.5, 2.5)$	1	0

Going through the above table, we can say that there are only 4 points counted as supported vectors $(x_{1,}\ x_{4,}\ x_{7,}\ x_9)$.

$$\omega = \sum_{i=1}^{N} \alpha_i x_i y_i$$

$$= 0.414(4 \quad 2.9) - 0.018(2.5 \quad 1) + 0.018(3.5 \quad 4) - 0.414(2 \quad 2.1)$$

$$= 0.414(4-2 \quad 2.9-2.1) + 0.018(3.5-2.5 \quad 4-1)$$

$$= 0.414(2 \quad 0.8) + 0.018(1 \quad 3)$$

$= (0.828 \quad 0.3312) + (0.018 \quad 0.054)$

$= (0.846 \quad 0.3852)$

$\boldsymbol{\omega} = (\mathbf{0.846} \quad \mathbf{0.3852})$

Now, we can calculate the bias as the average of the bias obtained from each supported vector by,

$$b_i = y_i - \omega^T x_i$$

$$b_1 = y_1 - \begin{pmatrix} 0.846 \\ 0.3852 \end{pmatrix} (4 \quad 2.9)$$

$$b_1 = 1 - (3.384 \quad 1.1136)$$

$$b_1 = (-2.384 \quad -0.1136)$$

$$\boldsymbol{b_1} = -\mathbf{2.4976}$$

$$b_4 = y_4 - \begin{pmatrix} 0.846 \\ 0.3852 \end{pmatrix} (2.5 \quad 1)$$

$$b_4 = -1 - (2.115 \quad 0.3852)$$

$$b_4 = (-3.115 \quad -1.3852)$$

$$\boldsymbol{b_4} = -\mathbf{4.5002}$$

$$b_7 = y_7 - \begin{pmatrix} 0.846 \\ 0.3852 \end{pmatrix} (3.5 \quad 4)$$

$$b_7 = 1 - (2.961 \quad 1.5408)$$

$$b_7 = (-1.961 \quad -0.5408)$$

$$\boldsymbol{b_7} = -\mathbf{2.5018}$$

$$b_9 = y_9 - \begin{pmatrix} 0.846 \\ 0.3852 \end{pmatrix} (2 \quad 2.1)$$

$$b_9 = -1 - (1.692 \quad 0.80892)$$

$$b_9 = (-2.692 \quad -1.80892)$$

$$\boldsymbol{b_9} = -\mathbf{4.50092}$$

$$b_{avg} = \frac{b_1 + b_4 + b_7 + b_9}{4}$$

$$= \frac{-2.4976 - 4.5002 - 2.5018 - 4.50092}{4}$$

$$= \frac{-14.00052}{4}$$

$$\boldsymbol{b_{avg} = -3.50013}$$

$\boldsymbol{\omega = (0.846 \quad 0.3852)}$

$\boldsymbol{b_{avg} = -3.50013}$

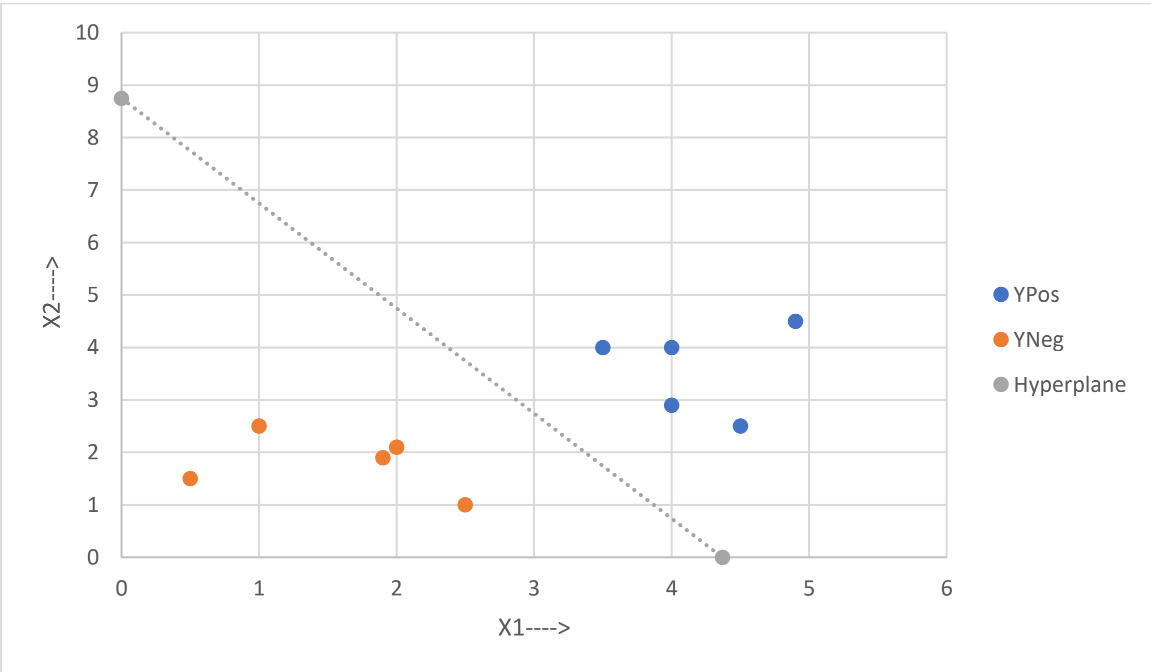
$$h(x) = \omega^T x + b_{avg}$$

$$= \begin{pmatrix} 0.8 \\ 0.4 \end{pmatrix} (x_1 \quad x_2) - 3.5$$

$$=> \boldsymbol{0.8x_1 + 0.4x_2 - 3.5 = 0}$$

For $x_1 = 0$, then $x_2 = \frac{3.5}{0.4} = 8.75$

For $x_2 = 0$, then $x_1 = \frac{3.5}{0.8} = 4.37$



(2)

Distance of x_6 $(1.9, 1.9)$ from hyperplane, $h(x)$: $0.8x_1 + 0.4x_2 - 3.5 = 0$,

Distance Formula $d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$

$$a = 0.8$$

$$b = 0.4$$

$$c = -3.5$$

$$x_0 = 1.9$$

$$y_0 = 1.9$$

$$d = \frac{|(0.8 * 1.9) + (0.4 * 1.9) - 3.5|}{\sqrt{(0.8)^2 + (0.4)^2}}$$

$$= \frac{|1.52 + 0.76 - 3.5|}{\sqrt{0.64 + 0.16}}$$

$$= \frac{1.22}{\sqrt{0.8}}$$

$$= \frac{1.22}{0.89}$$

$$\boldsymbol{d = 1.37}$$

No x_6 is not within the margin classifier.

(3)

$$z = (3,3)^T$$

$$h(x): 0.8x_1 + 0.4x_2 - 3.5 = 0$$

$$=> 0.8 * 3 + 0.4 * 3 - 3.5 = 0$$

$$=> 0.24 + 0.12 - 3.5 = 0$$

$$=> -3.14 \neq 0$$

As point z doesn't satisfies the $h(x)$ equation, so we can say it doesn't lie on the hyperplane.

$$f(x) = \omega^T x + b$$

$$= \begin{pmatrix} 0.8 \\ 0.4 \end{pmatrix} (3 \quad 3) - 3.5$$

$$= (0.24 \quad 0.12) - 3.5$$

$$= (0.24 - 3.5 \quad 0.12 - 3.5)$$

$$= (-3.26 \quad -3.38)$$

$$f(x) = -6.64$$

For classifying new observations,

$$\hat{y} = sign\big(f(x)\big)$$

$$\boldsymbol{\hat{y} = negative}$$

As, it's sign is showing negative.

Therefore, we can say that point $z = (3,3)^T$ belongs to negative class.