Sofia Baltzi 31/3/20

1(a)

(i) Since the classes are equiprapable we have that:

$$P(\omega_1)=P(\omega_2)=P(\omega_3)=1/3$$

For ω_1: $p(x|\omega_1)\cdot P(\omega_1)=1/3\cdot 1/3 = 1/9 = 0.1$, $x\in(3,4)\cup(6,8)$

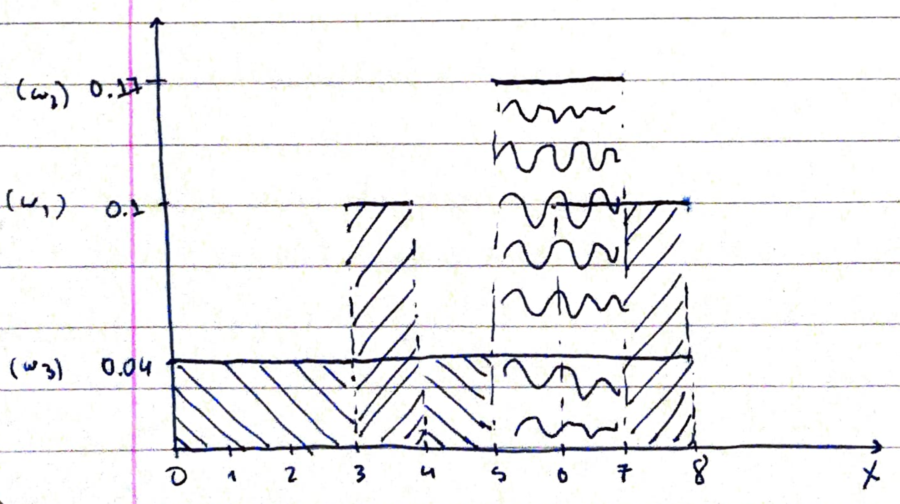
For ω_2: $p(x|\omega_2)\cdot P(\omega_2)=1/2\cdot 1/3 = 1/6 = 0.17$ $x\in(5,7)$

For ω_3: $p(x|\omega_3)\cdot P(\omega_3)=1/8\cdot 1/3 = 1/24 = 0.04$ $x\in(0,8)$

In order to decide in favour of 1 class over another we will use the Bayes rule. For classes i,j if $P(x|\omega_i)\cdot P(\omega_i) > P(x|\omega_j)P(\omega_j)$, the classification is in favour of class i. Therefore:

- (0,3): classification in favour of ω_3
- (3,4): classification in favour of ω_1
- (4,5): — " — ω_3
- (5,7): — " — ω_2
- (7,8): — " — ω_1

In the following diagram the horizontal lines are the $P(x|\omega_i)P(\omega_i)$ and the drawn areas the decision regions for the classes.



Decision regions for $\omega_1, \omega_2, \omega_3$:

$R_1 = \{3,4\} \cup \{7,8\}$: ////

$R_2 = \{5,7\}$: ~~

$R_3 = \{0,3\} \cup \{4,5\}$: \\\\

1

(ii) Having in mind that the classes are equiprapuble the propability of error is:

$$P_e = \frac{1}{3}\left(\int_{R_1} (p(x|\omega_2) + p(x|\omega_3))dx + \int_{R_2} (p(x|\omega_1) + p(x|\omega_3))dx + \int_{R_3} (p(x|\omega_1) + p(x|\omega_2))dx \right)$$

$$= \frac{1}{3}\left(\int_3^4 (p(x|\omega_2)+p(x|\omega_3))dx + \int_7^8 (p(x|\omega_2)+p(x|\omega_3))dx + \int_5^7 (p(x|\omega_1)+p(x|\omega_3))dx + \right.$$

$$\left. \int_0^3 (p(x|\omega_1)+p(x|\omega_2))dx + \int_4^5 (p(x|\omega_1)+p(x|\omega_2))dx \right) =$$

$$= \frac{1}{3}\left(\int_3^4 \left(0+\frac{1}{8}\right)dx + \int_7^8 \left(0+\frac{1}{8}\right)dx + \int_5^6 \left(0+\frac{1}{8}\right)dx + \int_6^7 \left(\frac{1}{3}+\frac{1}{8}\right)dx + \right.$$

$$\left. \int_0^3 (0+0)dx + \int_4^5 (0+0)dx \right) =$$

$$= \frac{1}{3}\left(\frac{1}{8}[x]_3^4 + \frac{1}{8}[x]_7^8 + \frac{1}{8}[x]_5^6 + \frac{11}{24}[x]_6^7 + 0 + 0 \right) =$$

$$= \frac{1}{3}\left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{11}{24} \right) = \frac{1}{3} \cdot \frac{20}{24} = 0{,}27$$

Propability of error is 0,27.

(iii) $x_1 = 2$ is classified to ω_3, $2 \in R_3$

$x_2 = 4.5$ is classified to ω_3, $4.5 \in R_3$

$x_3 = 5.5$ is classified to ω_2, $5.5 \in R_2$

1(b)(i) $p(x) = 2\theta x e^{-\theta x^2} u(x)$ (1)

Assuming statistical independence among the elements of X, the log-likelihood function of θ is:

$$L(\theta) = \ln p(X;\theta) = \ln\left(\prod_{i=1}^{N} p(x_i;\theta)\right) = \sum_{i=1}^{N} \ln(p(x_i;\theta))$$

Utilizing (1) we have that:

$$L(\theta) = \sum_{i=1}^{N} \ln(2\theta x e^{-\theta x^2}) = \sum_{i=1}^{N} \left[\overset{\ln(2)+\ln(\theta)}{\cancel{\ln(2\theta)}} + \ln(x) + \ln(e^{-\theta x^2})\right] =$$

$$= \sum_{i=1}^{N}\left[\ln(2\theta) + \ln(x_i) - \theta x_i^2\right] = \overset{N\ln(2)+N\ln(\theta)}{\cancel{N\ln(2\theta)}} + \sum_{i=1}^{N} \ln(x_i) - \theta \sum_{i=1}^{N} x_i^2$$

Maximum-likelihood estimate of θ is obtained by taking the gradient of the L and equating to 0.

$$\frac{\partial L(\theta)}{\partial \theta} = 0 \Rightarrow \frac{N}{\theta_{ML}} - \sum_{i=1}^{N} x_i^2 = 0 \Rightarrow \theta_{ML} = \frac{N}{\sum_{i=1}^{N} x_i^2}$$

(ii) $X_1 = \{0, 0.5, 1, 1.5, 2\}$, $N_1 = 5$
$X_2 = \{3, 3.5, 4, 4.5, 5\}$, $N_2 = 5$

I $\theta_{ML}^{1} = \dfrac{5}{0+0.25+1+2.25+4} = 0.67$, $p_1(x;\theta) = 1.34\, x \cdot e^{-0.67x^2}$

$\theta_{ML}^{2} = \dfrac{5}{9+12.25+16+20.25+25} = 0.06$, $p_2(x;\theta) = 0.12\, x \cdot e^{-0.06x^2}$

II For the a-priori $P(\omega_i) = \frac{N_i}{N}$, $N=10$, $i=1,2$
$P(\omega_1) = 1/2$, $P(\omega_2) = \frac{1}{2}$

According to Bayes rule, we decide in favour of ω_1 when
$P(\omega_1) \cdot p(x|\omega_1) > P(\omega_2)\, p(x|\omega_2)$ $\overset{P(\omega_1)=P(\omega_2)}{\Rightarrow}$

$$\Rightarrow 1.34x \cdot e^{-0.67x^2} > 0.12xe^{-0.06x^2} \overset{x>0}{\Rightarrow}$$

$$1.34\, e^{-0.67x^2} > 0.12\, e^{-0.06x^2} \Rightarrow$$

$$\ln(1.34) - 0.67x^2 > \ln(0.12) - 0.06x^2 \Rightarrow$$

$$0.29 - 0.67x^2 > -2.12 - 0.06x^2 \Rightarrow$$

$$2.41 > 0.61x^2 \Rightarrow$$

$$x^2 < 3.95 \Rightarrow$$

$$-1.98 < x < 1.99$$

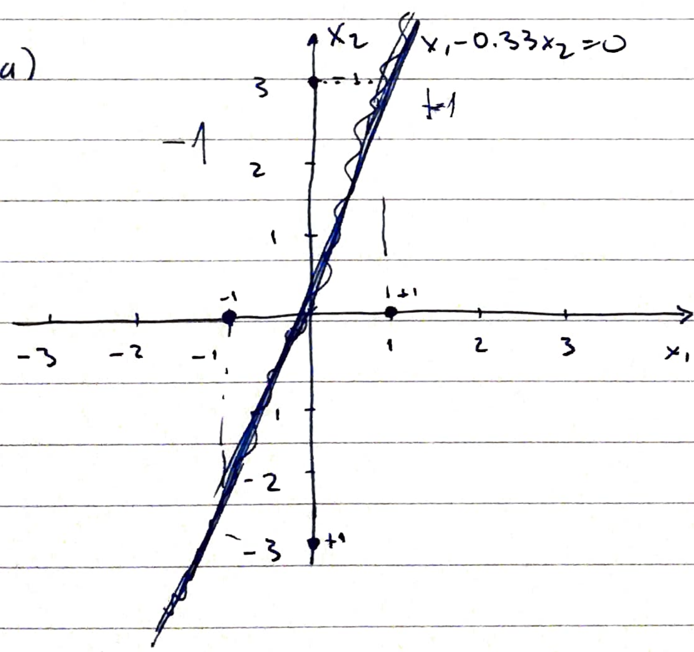
Thus, we classify to ω_1 when $x \in (-1.99, 1.99)$ and to ω_2 otherwise

We have made the assumption that $x>0$ thus the regions become: ω_1: $x \in (0, 1.99)$

ω_2: $x \in (1.99, +\infty)$

ω_1 or ω_2 if $x = 1.99$

2(a)



We will utilize the LS criterion wher $\theta = (X^T X)^{-1} X^T y$

θ is the parameter vector of the line that separates the classes.

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -3 \\ 1 & -1 & 0 \\ 1 & 0 & 3 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}$$

$$\bullet\ X^T X = \overset{3\times 4}{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & -3 & 0 & 3 \end{bmatrix}} \overset{4\times 3}{\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -3 \\ 1 & -1 & 0 \\ 1 & 0 & 3 \end{bmatrix}} =$$

$$= \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 18 \end{bmatrix}$$

$$\bullet\ (X^T X)^{-1} = \frac{1}{|X^T X|} \cdot \operatorname{adj}(X^T X)$$

$$\det(X^T X) = 4 \begin{vmatrix} 2 & 0 \\ 0 & 18 \end{vmatrix} = 144$$

$$\operatorname{adj}(X^T X) = \begin{vmatrix} 36 & 0 & 0 \\ 0 & 72 & 0 \\ 0 & 0 & 8 \end{vmatrix}$$

$$X^T X^{-1} = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.056 \end{bmatrix}$$

$$\bullet\ (X^T X)^{-1} \cdot X^T = \begin{bmatrix} 0.25 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.056 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & -3 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0 & -0.5 & 0 \\ 0 & -0.17 & 0 & 0.17 \end{bmatrix}$$

$$\bullet\ (X^T X)^{-1} X^T y = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.5 & 0 & -0.5 & 0 \\ 0 & -0.17 & 0 & 0.17 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ -0.33 \end{bmatrix}$$

$$\theta = [0 \quad 1 \quad -0.33]^T$$

The line that separates the two classes is:

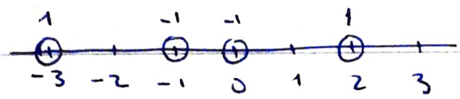
$$0 + x_1 - 0.33 x_2 = 0 \Rightarrow x_1 - 0.33 x_2 = 0$$

(ii) $[5, 1]^T$: $5 - 0.33 > 0 \Rightarrow$ class $+1$

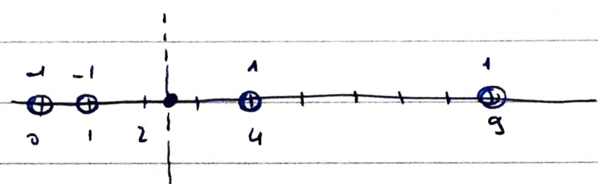
$[1, 5]^T$: $1 - 5 \cdot 0.33 < 0 \Rightarrow$ class -1

(iii) SVM would give the exact same line.
This line is paralel to the two lines that pass from the two points of each class and ~~to the distances~~ ~~at~~ passes in the midle of these two lines.
~~therefore~~

(b)



with $\phi(x)=x^2$ the points become

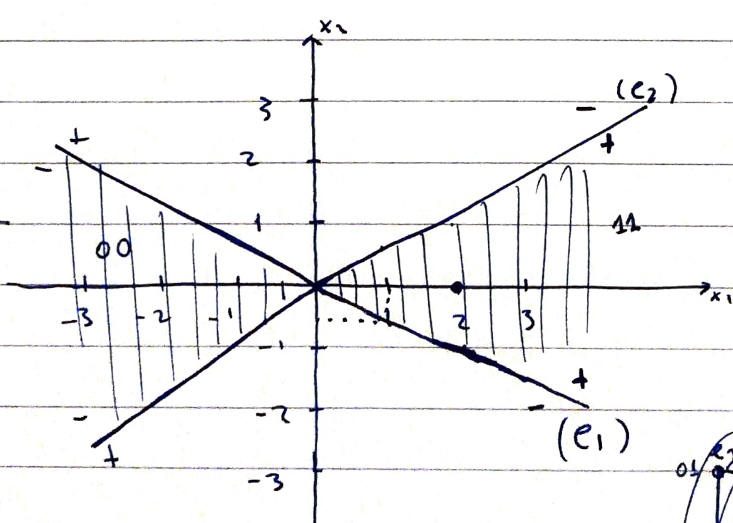


Therefore, it is linearly separable. The ~~line~~ point $x=2.5$ separates the two classes.

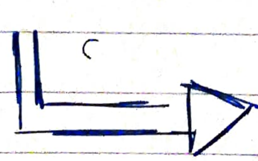
3(a) $x_1 + 2x_2 = 0$ (e_1)

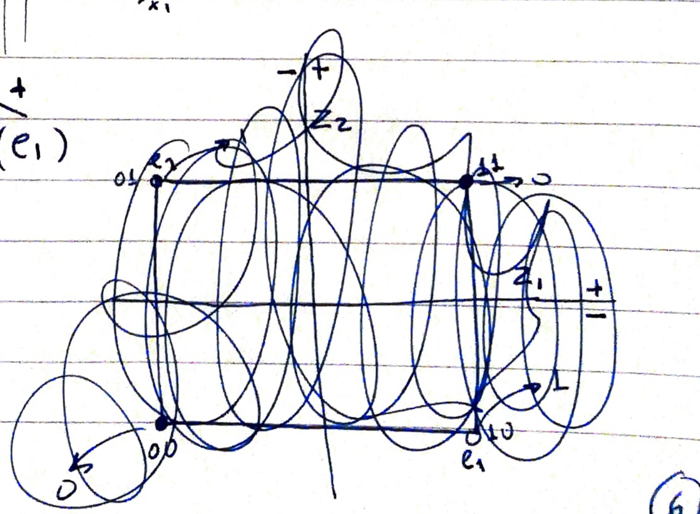
$x_1 - 2x_2 = 0$ (e_2)

fig1



class 0 : |||
class 1 :





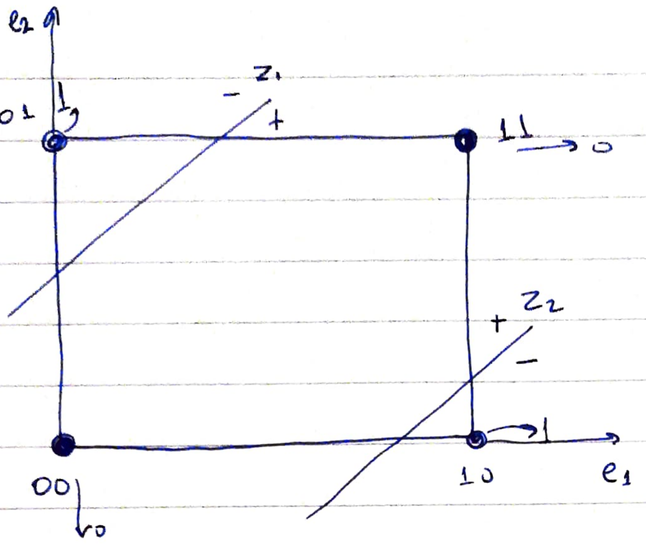


fig2

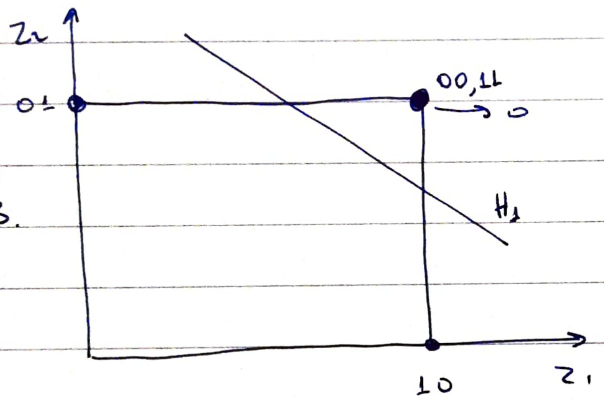


fig3.

- As we can see on (fig 1) the two classes are not linearly separable. Since they are separated with two lines, the 1st layer of the Neural Network will have two nodes with weights (1, 2) and (1, -2) respectively and biases equal to 0.

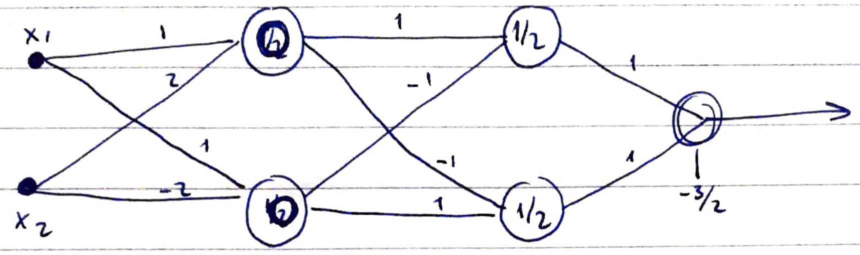
- Moving to ~~the second layer~~, (fig 2) the classes are still non-linearly separable. Therefore, we need another layer of two nodes (since the two classes are again separated by two lines).
~~The lines, intuitively are [illegible]~~

z_1: passes through the points $(0, \frac{1}{2})$, $(\frac{1}{2}, 1)$ therefore
$x_1 - x_2 + \frac{1}{2} = 0$: weights $(1, -1)$ bias $\frac{1}{2}$

z_2: passes through the points $(\frac{1}{2}, 0)$, $(1, \frac{1}{2})$ therefore
$-x_1 + x_2 + \frac{1}{2} = 0$: weights $(-1, 1)$ bias $\frac{1}{2}$

- Finally, the classes are linearly separable, by the line that passes through the points $(1, 1/2)$, $(1/2, 1)$

$$H_1 : x_1 + x_2 - 3/2 = 0 \quad : \text{weights } (1,1) \text{ bias } -3/2$$



(b) (i) From the 5 nearest neighbors, 3 belong to class ω_1 and 2 to class ω_2. Therefore, we classify the point to class 1.

(ii) For the probabilities $P(\omega_1)$, $P(\omega_2)$: They are equal since classes are equiprobable

$P(\omega_1) = \frac{N_{\omega_1}}{N} = \frac{18}{27}$

$P(\omega_1) = P(\omega_2) = 1/2$

$P(\omega_2) = \frac{N_{\omega_2}}{N} = \frac{9}{27}$

$N_1 = 18$, $N_2 = 9$

For ω_1 the circle that has the smallest radius and contains 3 neighbors has radius $r_1 = 1$ and $V_1 = \pi$ (πr_1^2)

For ω_2 the circle that has the smallest radius and contains 3 neighbors has radius $r_2 = \sqrt{5}$ and $V_2 = 5\pi$ (πr_2^2)

$$P(x|\omega_1) = \frac{k}{N_1 V_1} = \frac{3}{18 \cdot \pi} = 0.053$$

$$P(x|\omega_2) = \frac{k}{N_2 V_1} = \frac{3}{9 \cdot 5\pi} = 0.021$$

Since $P(x|\omega_1) \cdot P(\omega_1) > P(x|\omega_2) P(\omega_2)$ and based on Bayes classification rule we classify the point to class ω_1

4(a) feature space is separated by line $x_1 = 1/2$, thus the root of the tree will have the condition $x_1 \leq 1/2$.

$x_1 \leq 1/2$: YES

The given feature space is separated by line $x_2 = 1/4$, thus the condition will be $x_2 \leq 1/4$.

$x_2 \leq 1/4$: YES classify to class ω_3

$x_2 \cancel{\leq} 1/4$: No classify to class ω_1

Splitting process stops here for this branch, since classes are perfectly separated

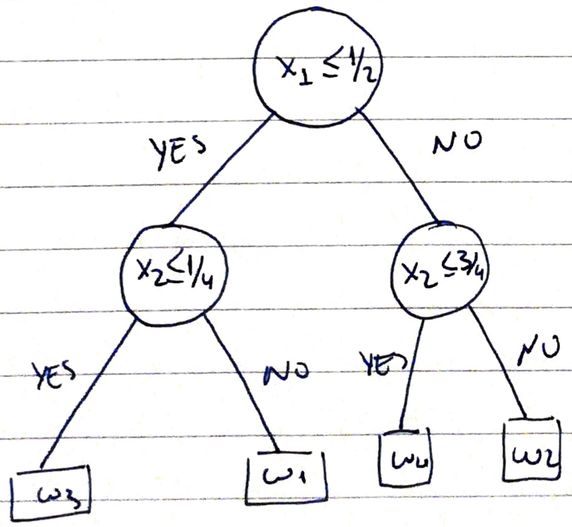
$x_1 \leq 1/2$: NO

The given feature space is separated by line $x_2 = 3/4$, thus the condition will be $x_2 \leq 3/4$.

$x_2 \leq 3/4$: YES classify to ω_4

$x_2 \leq 3/4$: NO classify to ω_2

Process stops here for this branch, since classes are perfectly separated.



(b) Since θ is unbiased estimator of θ we have that:

$E[\theta] = \theta_0$

We need θ' to be unbiased too, therefore $E[\theta'] = \theta_0 \Rightarrow$

$\Rightarrow E[(a^2-3)\theta] = (a^3-3)E[\theta] = (a^2-3)\theta_0$

$E[\theta'] = \theta_0 \Rightarrow (a^2-3)\theta_0 = \theta_0 \Rightarrow a^2-3 = 1 \Rightarrow$

$\Rightarrow a^2 = 4 \Rightarrow a = \pm 2$

For $a = \pm 2$ θ' is unbiased estimator of θ.

(c) $f(x) = x_1^2 + 4x_2 + 6$

~~$y_1 = 4 + 4 \cdot 3 + 6 \Rightarrow y_1 = 22$~~

~~$y_2 = 1 + 4 \cdot 1 + 6 \Rightarrow y_2 = 11$~~

$y_1 = 4 + 4 \cdot 3 + 6 \Rightarrow y = 22$

$y_2 = 1 + 4 \cdot 1 + 6 \Rightarrow y_2 = 11$

(d) (γ) both x_1 and x_2

None of the x_1, x_2 alone descriminate the classes, since the projection on the axes is the same for the two classes (no matter which axis you choose to project on)

Therefore, we need both axes x_1, x_2 to separate the classes.

(f) $x_1 = [1, 0]^T$, $x_2 = [3, 0]^T$, $x_3 = [1, 10]^T$, $x_4 = [3, 10]^T$

1st it) $\theta_1(0) = [4, 10]^T$ $\theta_2(0) = [4, 0]^T$

x_1	$109 (= (4-1)^2 + (10-0)^2)$	9	class ω_2
x_2	101	1	class ω_2
x_3	9	109	class ω_1
x_4	1	101	class ω_1

We computed the squared Euclidean distances and classified each point to a class by the rule $\min_{j=1,2} \|x_i - \theta_j(0)\|^2$ $i = 1, \ldots 4$

For the new θ:

$$\theta_1(1) = \frac{1}{2}\left(\begin{bmatrix}1\\10\end{bmatrix} + \begin{bmatrix}3\\10\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}4\\20\end{bmatrix} = \begin{bmatrix}2\\10\end{bmatrix}$$

$$\theta_2(1) = \frac{1}{2}\left(\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}3\\0\end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix}4\\0\end{bmatrix} = \begin{bmatrix}2\\0\end{bmatrix}$$

2nd it) $\theta_1(1) = [2 \ 10]^T$ $\theta_2(2) = [2 \ 0]^T$

x_1	101	1	class ω_2
x_2	101	1	class ω_2
x_3	1	101	class ω_1
x_4	1	101	class ω_1

$$\theta_1(2) = \frac{1}{2}\left(\begin{bmatrix}4\\20\end{bmatrix}\right) = \begin{bmatrix}2\\10\end{bmatrix}$$

$$\theta_2(2) = \frac{1}{2}\left(\begin{bmatrix}4\\0\end{bmatrix}\right) = \begin{bmatrix}2\\0\end{bmatrix}$$

Since for two iterations in a row the representatives do not change (classification does not change either for the x's) we stop the process.

In conclusion, x_1, x_2 are classified to ω_2 with $\theta_2 = [2 \ 0]^T$ and x_3, x_4 are classified to ω_1 with $\theta_1 = [2 \ 10]^T$

(e) PCA does not necessarily retain class separability because its is possible that ~~bye eliminating projection~~ projection on an ~~the PCA~~ axis does not separate the classes e.g.