Homework 5 - Written Answer Key

<u>Question 1</u>:

(Question)

Consider the following training data,

x_1	x_2	y
0	0	-1
2	2	-1
2	0	1
1	1.5	-1
3	0.5	1

(a) Plot the training data, and the hyperplane given by: $\mathbf{w}^T = [3, -3]$ and $w_0 = -3$

(b) Are the points linearly separable?

(c) For $\mathbf{w}$ and w_0 given above:

   i. compute the *functional* margin for each training example, and show the functional margin with respect to the set of training examples

   ii. compute the *geometric* margin for each training example, and show the geometric margin with respect to set of training examples

   iii. compute the *canonical weights* with respect to the training examples

(d) Identify which of the training examples are support vectors

(e) If we add the point $x = (1, 3)^T$ and $y = -1$ to the training data, does the margin change? Does separating hyperplane change? Do the support vectors change?

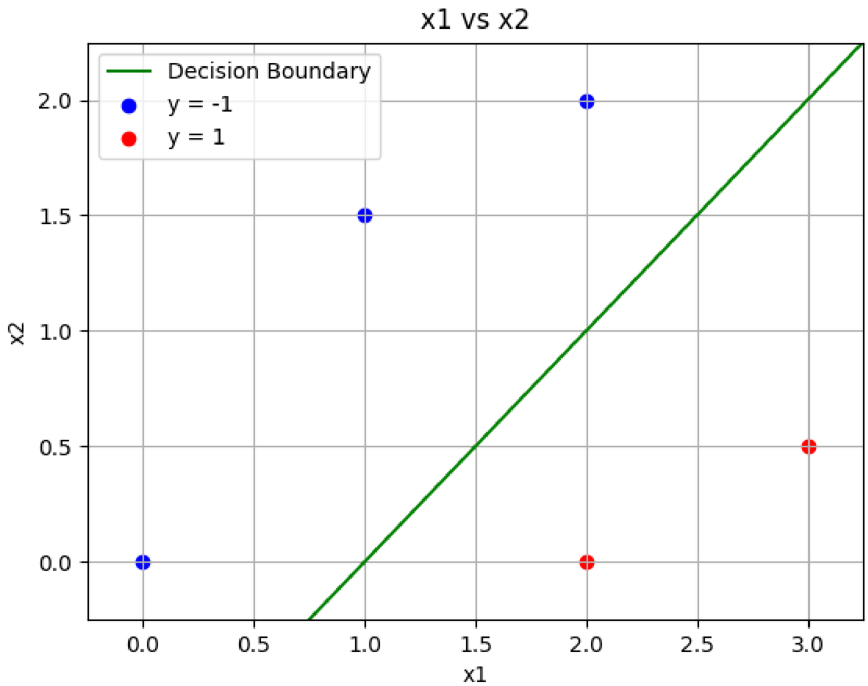
(f) If we remove the point $(1, 1.5)^T$ does the margin change? Does the separating hyperplane change?

(g) If we remove the point $(0, 0)^T$ does the margin change? Does the separating hyperplane change?

(h) Specify a *constrained optimization* to find a hyperplane that separates the training examples above where the separating hyperplane has the largest possible margin.[1] Use the numbers given in the problem.

(Answer(s))

1.



2.

The points are linearly separable (as shown by the decision boundary).

3.

$$\text{Functional margin formula: } \gamma^{(i)} = y^{(i)}(w^T x^{(i)} + w_0)$$

$$\gamma^{(1)} = (-1)(\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + -3)) = 3$$

$$\gamma^{(2)} = (-1)(\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + -3)) = 3$$

$$\gamma^{(3)} = (1)(\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} + -3)) = 3$$

$$\gamma^{(4)} = (-1)(\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 1.5 \end{bmatrix} + -3)) = 4.5$$

$$\gamma^{(5)} = (1)(\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 3 \\ 0.5 \end{bmatrix} + -3)) = 4.5$$

$$\text{Functional margin w.r.t. training examples } = \min(\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}, \gamma^{(4)}, \gamma^{(5)}) = 3$$

$$\text{Geometric margin formula: } \gamma^{(i)} = \frac{y^{(i)}(w^T x^{(i)} + w_0)}{||w||_2}$$

$$||w||_2 = \sqrt{3^2 + 3^2} = \sqrt{18}$$

$$\gamma^{(1)} = \frac{3}{\sqrt{18}} = \frac{1}{\sqrt{2}}$$

$$\gamma^{(2)} = \frac{3}{\sqrt{18}} = \frac{1}{\sqrt{2}}$$

$$\gamma^{(3)} = \frac{3}{\sqrt{18}} = \frac{1}{\sqrt{2}}$$

$$\gamma^{(4)} = \frac{4.5}{\sqrt{18}} = \frac{3}{\sqrt{8}}$$

$$\gamma^{(5)} = \frac{4.5}{\sqrt{18}} = \frac{3}{\sqrt{8}}$$

$$\text{Geometric margin w.r.t. training examples} = \min(\gamma^{(1)}, \gamma^{(2)}, \gamma^{(3)}, \gamma^{(4)}, \gamma^{(5)}) = \frac{1}{\sqrt{2}}$$

The functional margin is 3, so the canonical weights are:

$$w_0 = \frac{-3}{3} = -1$$

$$w = \frac{1}{3}\begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

4.

The points $x^{(1)}, x^{(2)}$, and $x^{(3)}$ ((0, 0), (2, 2), and (2, 0), respectively) are the support vectors because they each have the smallest functional margin.

5.

If this point is added, the margin, separating hyperplane, and support vectors do not change.

6.

If this point is removed, the margin and separating hyperplane do not change.

7.

If this point is removed, the margin and separating hyperplane will change.

8.

$$\min(||w||_2^2)$$

$$y^{(i)}(w_0 + w^T x^{(i)}) \geq 1 \text{ for i} = 1, \ldots, N$$

$$(-1)((-3) + \begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix}) \geq 1$$

$$(-1)((-3) + \begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}) \geq 1$$

$$(1)((-3) + \begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix}) \geq 1$$

$$(-1)((-3) + \begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 1.5 \end{bmatrix}) \geq 1$$

$$(1)((-3) + \begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 3 \\ 0.5 \end{bmatrix}) \geq 1$$

Question 2:

(Question)

Would the following constrained optimizations create the same decision boundary? Justify your answer.

$\texttt{max}\ \gamma$
Subject to:
$-1(5.1w_1 + 3.5w_2 + 1.4w_3 + 0.2w_4 + w_0) \geq \gamma$
$-1(4.9w_1 + 3.0w_2 + 1.4w_3 + 0.2w_4 + w_0) \geq \gamma$
$-1(4.7w_1 + 3.2w_2 + 1.3w_3 + 0.2w_4 + w_0) \geq \gamma$
$-1(4.6w_1 + 3.1w_2 + 1.5w_3 + 0.2w_4 + w_0) \geq \gamma$
$1(7.0w_1 + 3.2w_2 + 4.7w_3 + 1.4w_4 + w_0) \geq \gamma$
$1(16.4w_1 + 3.2w_2 + 4.5w_3 + 1.5w_4 + w_0) \geq \gamma$
$1(6.9w_1 + 3.1w_2 + 4.9w_3 + 1.5w_4 + w_0) \geq \gamma$
$1(.5w_1 + 2.3w_2 + 4.0w_3 + 1.3w_4 + w_0) \geq \gamma$
$||\mathbf{w}||_2 = 1$

$\texttt{min}\ ||\mathbf{w}||_2^2$
Subject to:
$-1(5.1w_1 + 3.5w_2 + 1.4w_3 + 0.2w_4 + w_0) \geq 1$
$-1(4.9w_1 + 3.0w_2 + 1.4w_3 + 0.2w_4 + w_0) \geq 1$
$-1(4.7w_1 + 3.2w_2 + 1.3w_3 + 0.2w_4 + w_0) \geq 1$
$-1(4.6w_1 + 3.1w_2 + 1.5w_3 + 0.2w_4 + w_0) \geq 1$
$1(7.0w_1 + 3.2w_2 + 4.7w_3 + 1.4w_4 + w_0) \geq 1$
$1(16.4w_1 + 3.2w_2 + 4.5w_3 + 1.5w_4 + w_0) \geq 1$
$1(6.9w_1 + 3.1w_2 + 4.9w_3 + 1.5w_4 + w_0) \geq 1$
$1(.5w_1 + 2.3w_2 + 4.0w_3 + 1.3w_4 + w_0) \geq 1$

(Answer(s))

The two constraints will create the same decision boundary.

$$\text{Maximizing } \gamma \rightarrow \text{Maximizing } \frac{\gamma}{1}$$

$$\text{Maximizing } \frac{\gamma}{1} \rightarrow \text{Maximizing } \frac{\gamma}{||w||_2}$$

$$\text{Maximizing } \frac{\gamma}{||w||_2} \rightarrow \text{Maximizing } \frac{1}{||\frac{w}{\gamma}||_2}$$

$$\text{Maximizing } \frac{1}{||\frac{w}{\gamma}||_2} \rightarrow \text{Minimizing } ||\frac{w}{\gamma}||_2$$

$$\text{Minimizing } ||\frac{w}{\gamma}||_2 \rightarrow \text{Minimizing } ||\frac{w}{\gamma}||_2^2$$

$$\text{Scale the rest of the constraints by a factor of } \frac{1}{\gamma} \text{ to finally obtain}$$

$$\text{Minimizing } ||\frac{w}{\gamma}||_2^2 \rightarrow \text{Minimizing } ||w||_2^2$$

Thus, the two constraints are the same.

Question 3:

(Question)

In class we discussed the soft-margin SVM:

$$\min \frac{1}{2}w^T w + C\sum_{i=1}^{n}\xi^{(i)}$$

$$\texttt{subject to } y^{(i)}(w^T x^{(i)} + w_0) \geq 1 - \xi^{(i)}, \texttt{ and } \xi^{(i)} \geq 0$$

(a) For a point $x^{(i)}$ if $\xi^{(i)} = 0$, what do we know about the where $x^{(i)}$ is wrt the margin. Is $x^{(i)}$ correctly classified by the hyperplane?

(b) For a point $x^{(i)}$ if $0 < \xi^{(i)} < 1$, what do we know about the where $x^{(i)}$ is wrt the margin. Is $x^{(i)}$ correctly classified by the hyperplane?

(c) For a point $x^{(i)}$ if $\xi^{(i)} > 1$, what do we know about the where $x^{(i)}$ is wrt the margin. Is $x^{(i)}$ correctly classified by the hyperplane?

(Answer(s))

1. If $\xi^{(i)} = 0$, we know that $x^{(i)}$ is correctly classified by the hyperplane. Additionally, $x^{(i)}$ can either be a support vector, or a point that is farther away from the support vector (still correctly classified).

2. If $0 < \xi^{(i)} < 1$, we know that $x^{(i)}$ is still correctly classified. However, $x^{(i)}$ must be located between the support vector boundary and the hyperplane.

3. If $\xi^{(i)} > 1$, $x^{(i)}$ must be incorrectly classified, so $x^{(i)}$ is somewhere on the wrong side of the hyperplane.

Question 4:

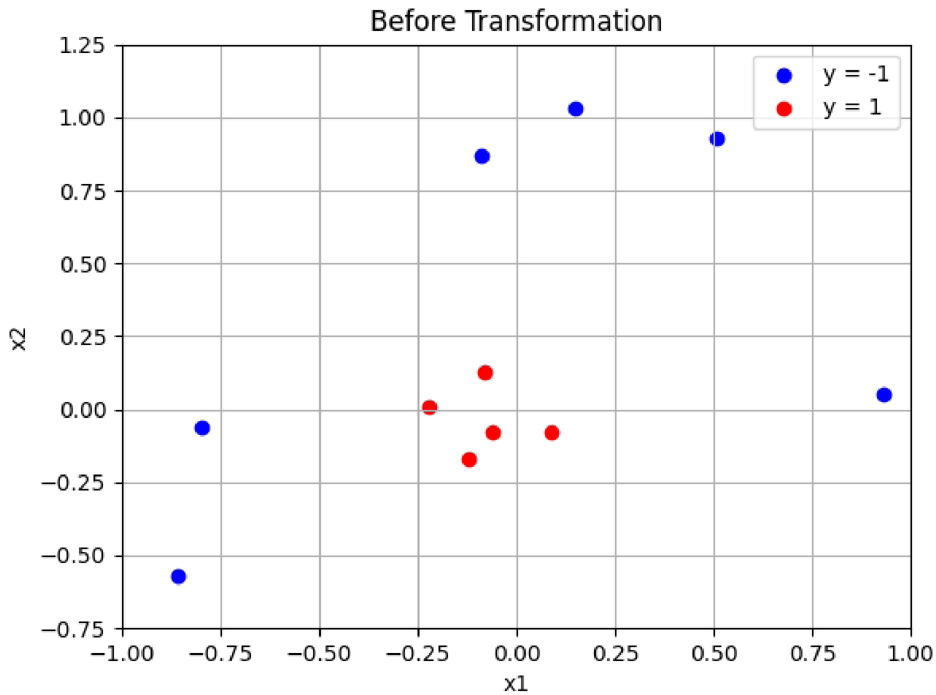
(Question)

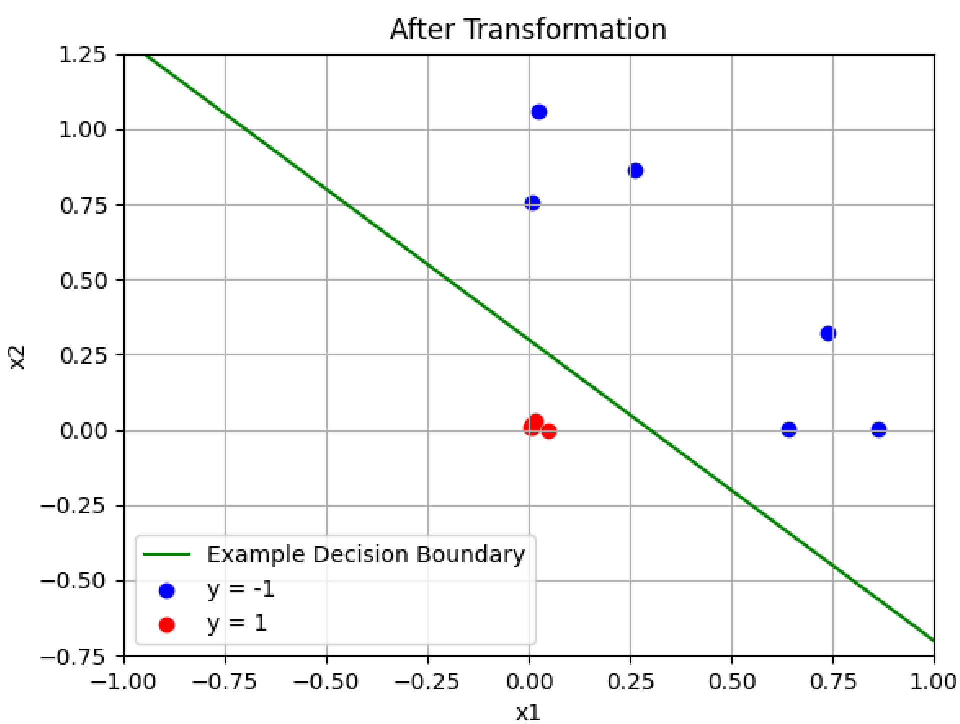
For the following non linearly separable points, find a transformation to make them linearly separable. Using matplotlib to plot the points before and after your transformation.

$x_1^{(i)}$	$x_2^{(i)}$	$y^{(i)}$
0.15	1.03	-1
-0.08	0.13	1
-0.80	-0.06	-1
-0.22	0.01	1
0.51	0.93	-1
-0.09	0.87	-1
0.09	-0.08	1
-0.12	-0.17	1
0.93	0.05	-1
-0.06	-0.08	1
-0.86	-0.57	-1

(Answer(s))

The transformation $\Phi(x^{(i)}) = \begin{bmatrix} (x_1^{(i)})^2 \\ (x_2^{(i)})^2 \end{bmatrix}$ will make the points linearly separable.


Before Transformation
y = -1
y = 1
x2
x1
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00


After Transformation
Example Decision Boundary
y = -1
y = 1
x2
x1
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00

Question 5:

(Question)

Consider the following training examples
$((1, 2.5)^T, 1), ((0, 0.75)^T, -1), ((1, 1)^T, -1), ((2, 2)^T, 1), ((3, 1)^T, 1), ((2, 3)^T, 1)$.

- After running the dual formalization, $\alpha = [0.9, 0, 1.4, 0, 0.5, 0]$. What is $\mathbf{w}$?
- What[2] is w_0?
- Predict the label of $(1, 0)^T$ and $(3, 3)^T$. Show your work.

(Answer(s))

1.

$$w = \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} x^{(i)}$$

$$w = (0.9)(1)(\begin{bmatrix} 1 \\ 2.5 \end{bmatrix}) + (0)(-1)(\begin{bmatrix} 0 \\ 0.75 \end{bmatrix}) + (1.4)(-1)(\begin{bmatrix} 1 \\ 1 \end{bmatrix}) + (0)(1)(\begin{bmatrix} 2 \\ 2 \end{bmatrix}) + (0.5)(1)(\begin{bmatrix} 3 \\ 1 \end{bmatrix}) + (0)(1)(\begin{bmatrix} 2 \\ 3 \end{bmatrix})$$

$$w = \begin{bmatrix} 1 \\ 1.35 \end{bmatrix}$$

2.

Use any of the points where $\alpha^{(i)} \neq 0$. This includes $i = 1, 3, 5$. In our calculations, we will use $x^{(3)}$

$$w_0 + w^T x = y$$

$$w_0 + \begin{bmatrix} 1 & 1.35 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = -1$$

$$w_0 = -3.35$$

3.

Predicting point $\begin{bmatrix}1\\0\end{bmatrix}$:

$$\hat{y} = \text{sign}(w_0 + w^T x)$$

$$\hat{y} = \text{sign}(-3.35 + \begin{bmatrix}1 & 1.35\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix})$$

$$\hat{y} = -1$$

Predicting point $\begin{bmatrix}3\\3\end{bmatrix}$:

$$\hat{y} = \text{sign}(-3.35 + \begin{bmatrix}1 & 1.35\end{bmatrix}\begin{bmatrix}3\\3\end{bmatrix})$$

$$\hat{y} = 1$$

Question 6:

(Question)

Given the following 5 training points:

i	$x_1^{(i)}$	$x_2^{(i)}$	$y^{(i)}$
1	0	0	1
2	-1	-1	1
3	1	1	-1
4	2	2	-1
5	0	-2	1

What would be the Lagrangian parameters $\alpha^{(i)}$ that would be found for a linear kernel hard margin SVM?

(a) $\alpha^{(1)} = -1$, $\alpha^{(4)} = -1$
(b) $\alpha^{(1)} = -1$, $\alpha^{(3)} = -1$
(c) $\alpha^{(1)} = 1$, $\alpha^{(4)} = 1$
(d) $\alpha^{(1)} = 1$, $\alpha^{(3)} = 1$
(e) $\alpha^{(1)} = 2$, $\alpha^{(4)} = 2$
(f) $\alpha^{(1)} = 2$, $\alpha^{(3)} = 2$

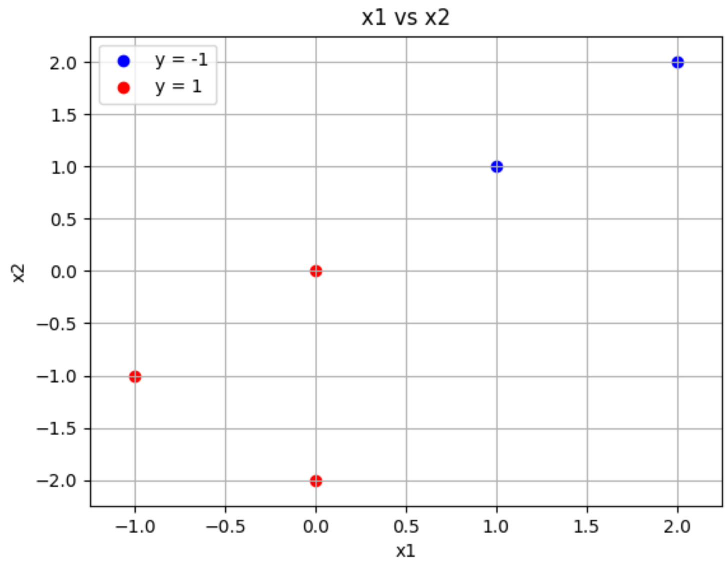
What would the value of w_0?

(a) $w_0 = 1$
(b) $w_0 = 0$
(c) $w_0 = -1$

(This question is modified from an outside source.)

Hint: Plot the points.

(Answer(s))



As seen in the graph above, the support vectors are going to be $x^{(1)}$ and $x^{(3)}$ (which are the points $(0,0)$ and $(1,1)$, respectively). This means that only $\alpha^{(1)}, \alpha^{(3)} \neq 0$. This rules out choices (a), (c), and (e). Additionally, Lagrange multipliers are positive, so (b) can be eliminated as well. We are now choosing between (d) and (f).

$$w = \sum_{i=1}^{N} \alpha^{(i)} y^{(i)} x^{(i)}$$

$$w = \alpha^{(1)} y^{(1)} x^{(1)} + \alpha^{(3)} y^{(3)} x^{(3)}$$

$$w = \alpha^{(1)}(1) \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \alpha^{(3)}(-1) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$w = \alpha^{(1)} \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \alpha^{(3)} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$y^{(1)}(w^T x^{(1)} + w_0) = (1)(w^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} + w_0)$$

However, in the case of a hard margin, support vectors have a functional margin of 1. This means that

$$1 = (1)(w^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} + w_0)$$

$$1 = (\alpha^{(1)} \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \alpha^{(3)} \begin{bmatrix} 1 \\ 1 \end{bmatrix})^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} + w_0$$

$$w_0 = 1$$

This means the answer to Q6.2 is (a).

Now that we have w_0, plug it back into the formula for $y^{(3)}$ to find $\alpha^{(3)}$

$$y^{(3)}(w^T x^{(3)} + w_0) = (-1)(w^T \begin{bmatrix} 0 \\ 0 \end{bmatrix} + w_0)$$

$$1 = (-1)((\alpha^{(1)} \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \alpha^{(3)} \begin{bmatrix} 1 \\ 1 \end{bmatrix})^T \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1)$$

$$\alpha^{(3)} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2$$

$$\alpha^{(3)} = 1$$

This means the answer to Q6.1 is (d)