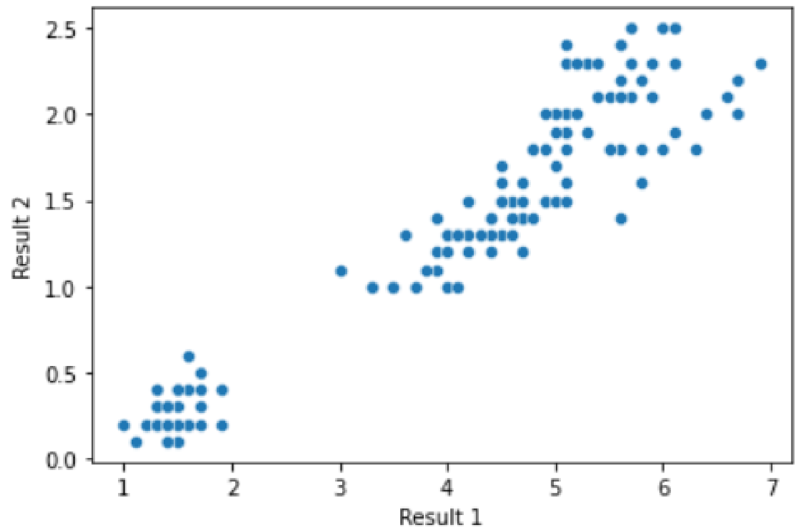
Assignment 4, CLL 788
Ansh Lodhi, 2019CH70161

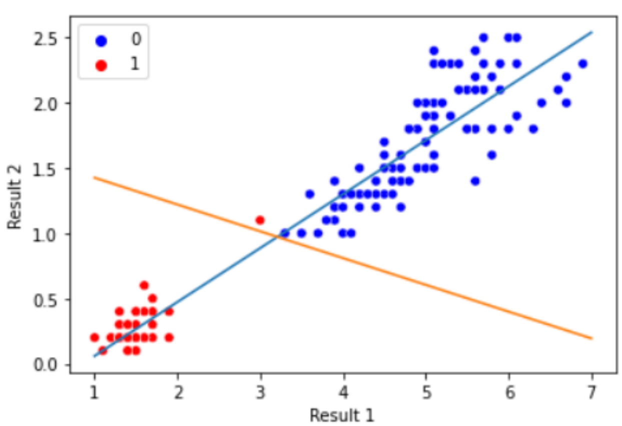
Answer 2 (a):

```
-----------------------Answer 2 (a): Data.xlsx ------------------------

<matplotlib.axes._subplots.AxesSubplot at 0x7fc507967ed0>
```



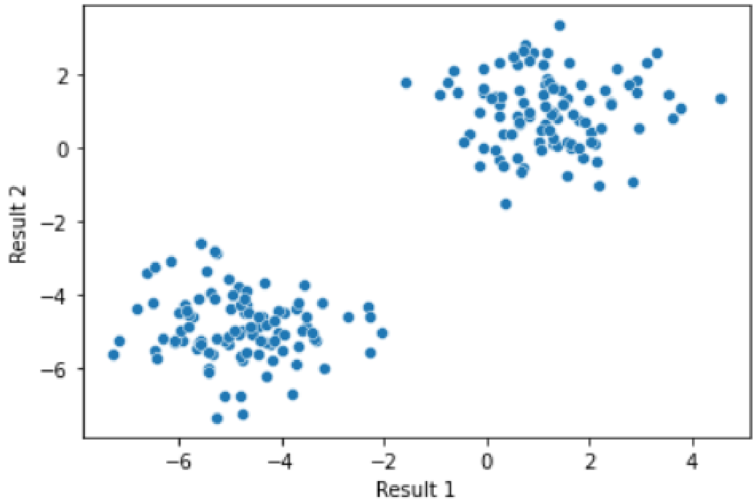
Answer 2 (b):

```
---------------- Answer 2 (b) : Data.xlsx --------------------------

Mean Values are as follows:
 [[4.92525253 1.68181818]
 [1.49215686 0.2627451 ]]

Orange line is the line passing from the midpoint of mean and is orthogonal to the line joing the two means
```



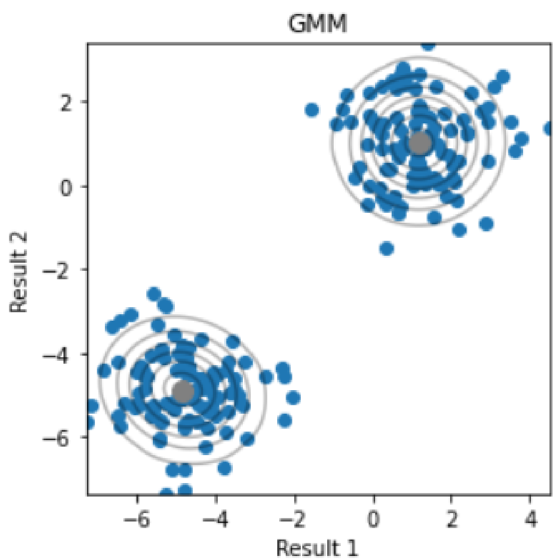
Answer 2 (c):

```
-----------------------Answer 2 (c): Data_GMM.xlsx ------------------------

<matplotlib.axes._subplots.AxesSubplot at 0x7fc5064c8550>
```



Answer 2 (d):

```
----------------- Answer 2 (d) : Data_GMM.xlsx -----------------------------

<Figure size 288x288 with 0 Axes>
```



Answer 1 :

Assignment - 4

ANSH LODHI
2019CHA0161

①

S. No	Var 1	Variable 2
1	-1.54	2.29
2	-0.44	2.34
3	-0.03	0.41
4	1.2	1.87
5	0.65	2.39
6	-4.67	-4.8
7	-3.37	-5.41
8	-3.93	-4.64
9	-4.78	-4.96
10	-4.12	-5.36

Assuming Variable 1 as x & Variable 2 as y.

∴ We can write each point as (x_i, y_i) $[1 \leq i \leq 10]$

∵ We have to divide these points into 2 clusters

∴ We assume 2 means : $\mu_1 (x_3, y_3)$ & $\mu_0 (x_6, y_6)$

S.No	Distance from μ_0	Distance from μ_1	Cluster
1	7.75	2.44	1
2	8.29	1.98	1
3	7.01	0	1
4	8.88	1.87	1
5	8.94	2.07	1
6	0	7.01	0
7	1.43	6.74	0
8	0.75	6.41	0
9	0.19	7.20	0
10	0.78	7.1	0

Calculating the distance of each point from both the assumed means μ_0 & μ_1 using formula —

$$\text{Distance} = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2}$$

$$= \sqrt{(\mu_x - x_i)^2 + (\mu_y - y_i)^2}$$

Assigning the cluster to each point on the basis of its distance from the assumed means. Points were assigned to the cluster which had less distance from mean.

Now, we assume the cluster assignment as constant and would shift the mean.

$\therefore$ for cluster 0:

x-co-ordinate of new mean

$$= \frac{-4.67 - 3.37 - 3.93 - 4.78 - 4.12}{5}$$

$$x = -4.174$$

y-coordinate of new mean

$$= \frac{-4.8 - 5.41 - 4.64 - 4.96 - 5.36}{5}$$

$$y = -5.034$$

Similarly, for cluster 1:

$$x\text{-coordinate} = \frac{-1.54 - 0.44 + 0.03 + 1.2 + 0.65}{5} = -0.02$$

$$y\text{-coordinate} = \frac{2.29 + 2.34 + 0.41 + 1.87 + 2.39}{5} = 1.86$$

$\mu_0 = (-4.174, -5.034)$ & $\mu_1 = (-0.02, 1.86)$

Table-3

S.No	Distance from μ_0	Distance from μ_1	Cluster
1	7.78	1.57	1
2	8.26	0.63	1
3	6.87	1.45	1
4	8.74	1.22	1
5	8.85	0.85	1
6	0.54	8.12	0
7	0.88	8.0	0
8	0.46	7.58	0
9	0.61	8.31	0
10	0.33	8.30	0

Now, we shift the mean

for cluster 0 : $X = \dfrac{-4.67-3.37-3.93-4.78-4.12}{5}$

$X = -4.174$

$Y = \dfrac{-4.8-5.41-4.64-4.96-5.36}{5}$

$Y = -5.034$

for cluster 1 : $X = \dfrac{-1.54-0.44+0.03+1.2+0.65}{5}$

$X = -0.02$

$Y = \dfrac{2.29+2.34+0.41+1.87+2.39}{5} = 1.86$

$\mu_0(-4.174, -5.034)$ & $\mu_1 = (-0.02, 1.86)$

The means are remaining same.

$\Rightarrow$ They have converged $\Rightarrow$ Table 3 shows correct classification.

Answer 3 :

(3).

Y_1	Y_2
2	1
3	4
5	0
7	6
9	2

First we make the data centered around 0 by subtracting the mean from both columns.

$$\text{Mean for } Y_1 = \frac{2+3+5+7+9}{5} = 5.2$$

$$\text{Mean of } Y_2 = \frac{1+4+0+6+2}{5} = 2.6$$

New Table :

Y_1 :	-3.2	-2.2	-0.2	1.8	3.8
Y_2 :	-1.6	1.4	-2.6	3.4	-0.6

Now we will calculate the covariance matrix S.

$\because$ We have 2-D data

$\therefore$ S has dimensions 2x2

$$S = \frac{1}{N} \sum_{i=1}^{N} (x_n - \bar{x})(x_n - \bar{x})^T$$

$$S_{11} = \frac{1}{5}\left[(-3.2)^2 + (-2.2)^2 + (-0.2)^2 + (1.8)^2 + (3.8)^2\right]$$

$$= \frac{1}{5}\left[10.24 + 4.84 + 0.04 + 3.24 + 14.44\right]$$

$$= 6.56$$

$$S_{22} = \frac{1}{5}\left[(-1.6)^2 + (1.4)^2 + (-2.6)^2 + (3.4)^2 + (-0.6)^2\right]$$

$$= 4.64$$

$$S_{12} = S_{21} = \frac{1}{5}\left[(-3.2)(-1.6) + (-2.2)(1.4) + (-0.2)(-2.6) + (1.8)(3.4) + (3.8)(-0.6)\right]$$

$$= \frac{1}{5}\left(5.12 - 3.08 + 0.52 + 6.12 - 2.28\right]$$

$$= 1.28$$

$$\therefore\ S = \begin{bmatrix} 6.56 & 1.28 \\ 1.28 & 4.64 \end{bmatrix}$$

λ : eigen value & U: eigen vector.

$$SU = \lambda U \Rightarrow$$

$$\Rightarrow (S - \lambda)U = 0$$

$$\left(\begin{bmatrix} 6.56 & 1.28 \\ 1.28 & 4.64 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) U = 0$$

$$\begin{bmatrix} 6.56-\lambda & 1.28 \\ 1.28 & 4.64-\lambda \end{bmatrix} U \; = 0$$

$$(6.56-\lambda)(4.64-\lambda) - (1.28)(1.28) = 0$$

$$30.4384 + \lambda^2 - 11.2\lambda - 1.6384 = 0$$

$$\lambda^2 - 11.2\lambda + 28.8 = 0$$

$$\boxed{\lambda = 4 \quad \& \quad \lambda = 7.2} \rightarrow \text{Eigen values.}$$

$$\begin{bmatrix} 6.56 & 1.28 \\ 1.28 & 4.64 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 7.2 \begin{bmatrix} x \\ y \end{bmatrix} \qquad (\text{for } \lambda = 7.2)$$

$$6.56x + 1.28y = 7.2x$$
$$1.28x + 4.64y = 7.2y$$

By solving the above 2 eqⁿs we get —

$$\boxed{x = 2y}$$

$$\begin{bmatrix} 6.56 & 1.28 \\ 1.28 & 4.64 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 4 \begin{bmatrix} x \\ y \end{bmatrix} ; \text{ for } (\lambda = 4)$$

$$6.56x + 1.28y = 4x$$
$$1.28x + 4.64y = 4y$$

By solving the above 2 eqⁿ we get:

$$-2x = y.$$

$\therefore$ Eigen vectors $u_1 = \begin{bmatrix} +2 \\ +1 \end{bmatrix}$; $u_2 = \begin{bmatrix} -1 \\ +2 \end{bmatrix}$

$$U = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \Rightarrow U = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

we can also transform u_1 & u_2 to unit vectors.

$$u_1 = \begin{bmatrix} 0.89 \\ 0.44 \end{bmatrix} \; \& \; u_2 = \begin{bmatrix} -0.44 \\ 0.89 \end{bmatrix}$$

$$U = \begin{bmatrix} 0.89 & -0.44 \\ 0.44 & 0.89 \end{bmatrix}$$

Transforming data to new co-ordinate space.

$$\begin{bmatrix} -3.2 & -1.6 \\ -2.2 & 1.4 \\ -0.2 & -2.6 \\ 1.8 & 3.4 \\ 3.8 & -0.6 \end{bmatrix} \cdot \begin{bmatrix} 0.89 & -0.44 \\ 0.44 & 0.89 \end{bmatrix} = \begin{bmatrix} -3.55 & -0.01 \\ -1.34 & 2.21 \\ -1.32 & -2.22 \\ 3.09 & 2.23 \\ 3.11 & -2.2 \end{bmatrix}$$

These are the new data values for PC1 & PC2

If we want to reduce the dimension of dataset then we can take the 1st column of dataset which will then become 1-D representation.

$$\boxed{Y: -3.55, -1.34, -1.32, 3.09, 3.11}$$