Name:- Aishwarya Vinod Menon

Machine Learning NetID: AXV220062

Homework-7

Question1

Part-1

$m = 4$, $k = 2$

1a.

1 (0,0)
2 (4,0)
3 (5,1)
4 (6,0)

given, $u_1 = (5,1)$ $u_2 = (6,0)$

Calculating distances.

For (0,0)

$d_1 = \sqrt{(0-5)^2+(0-1)^2} = \sqrt{26} = 5.099$

$d_2 = \sqrt{(0-6)^2+(0-0)^2} = 6$

For (4,0)

$d_1 = \sqrt{(4-5)^2+(0-1)^2} = \sqrt{2} = 1.414$

$d_2 = \sqrt{(6-4)^2+(0-0)^2} = 2$

For (5,1)

$d_1 = 0$

$d_2 = \sqrt{(6-5)^2+(1-0)^2} = \sqrt{2} = 1.414$

For (6,0)

$d_1 = \sqrt{(6-5)^2+(0-1)^2} = \sqrt{2} = 1.414$

		Arranged to cluster.
1	(0,0)	u_1
2	(4,0)	u_1
3	(5,1)	u_1
4	(6,0)	u_2

2. Recalculating centroid

$$u_1 = \left(\frac{(0+4+5)}{3}, \frac{(0+1+0)}{3}\right) = \left(\frac{9}{3}, \frac{1}{3}\right) = \left(3, \frac{1}{3}\right)$$

$u_2 = (6,0)$

Rearranging points to the nearest centroid

For (0,0)

$$d_1 = \sqrt{(3-0)^2 + \left(\frac{1}{3}-0\right)^2} = \sqrt{\frac{82}{9}} = 3.018$$

$$d_2 = \sqrt{(6-0)^2 + (0-0)^2} = 6$$

For (4,0)

$$d_1 = \sqrt{(3-4)^2 + \left(\frac{1}{3}-0\right)^2} = \sqrt{\frac{9+1}{9}} = \sqrt{\frac{10}{9}} = 1.05$$

$$d_2 = \sqrt{(6-4)^2 + (0-0)^2} = 2$$

For (5,1)

$$d_1 = \sqrt{(3-5)^2 + \left(\frac{1}{3}-1\right)^2} = \sqrt{\frac{40}{9}} = 2.108$$

$$d_2 = \sqrt{(5-6)^2 + (1-0)^2} = \sqrt{2} = 1.414$$

Points		Cluster
1	(0,0)	u_1
2	(4,0)	u_1
3	(5,1)	u_2
4	(6,0)	u_2

3. Second Iteration.

Recalculating centroids.

$u_1 = \left(\frac{(0+4)}{2}, \frac{0+0}{2}\right) = (2,0)$

$u_2 = \left(\frac{5+6}{2}, \frac{1+0}{2}\right) = \left(\frac{11}{2}, \frac{1}{2}\right)$

Calculating distances from new centroids

For (0,0)

$d_1 = \sqrt{(2-0)^2 + (0-0)^2} = 2$

$d_2 = \sqrt{\left(\frac{11}{2}-0\right)^2 + \left(\frac{1}{2}-0\right)^2} = \sqrt{\frac{122}{4}} = 5.522$

For (4,0)

$d_1 = \sqrt{(4-2)^2 + (0-0)^2} = 2$

$d_2 = \sqrt{\left(\frac{11}{2}-4\right)^2 + \left(\frac{1}{2}-0\right)^2} = 1.557$

For (5,1)

$d_1 = \sqrt{(5-2)^2 + (1-0)^2} = \sqrt{10} = 3.162$

$d_2 = \sqrt{(5.5-5)^2 + (0.5-1)^2} = 0.7071$

For (6,0)

$u_1 = \sqrt{(6-2)^2 + (0-0)^2} = 4$

$u_2 = \sqrt{(6-5.5)^2 + (0-0.5)^2} = \sqrt{0.5} = 0.71$

	Points	cluster
1	(0,0)	u_1
2	(4,0)	u_2
3	(5,1)	u_2
4	(6,0)	u_2

4. Third iteration

Recalculating centroids

$u_1 = (0,0)$

$u_2 = \left(\frac{4+5+6}{3}, \frac{1}{3}\right) = \left(5, \frac{1}{3}\right)$

Calculating distances.

For (6,0)

$d_1 = \sqrt{(0-0)^2 + (0-0)^2} = (0,0)$

$d_2 = \sqrt{(5-0)^2 + \left(\frac{1}{3}-0\right)^2} = \sqrt{25 + \frac{1}{9}} = \sqrt{\frac{34}{9}} = 5.01$

For (4,0)

$d_1 = \sqrt{(4-0)^2 + (0-0)^2} = 4$

$$d_2 = \sqrt{(5-4)^2 + \left(\frac{1}{3}-0\right)^2} = \sqrt{1+\frac{1}{9}} = \sqrt{\frac{10}{9}} = 1.05$$

For (5, 1)

$$d_1 = \sqrt{(5-0)^2 + (1-0)^2} = \sqrt{26} = 5.10$$

$$d_2 = \sqrt{(5-5)^2 + \left(1-\frac{1}{3}\right)^2} = \sqrt{\left(\frac{2}{3}\right)^2} = 0.67$$

For (6, 0)

$$d_1 = 6$$

$$d_2 = \sqrt{(6-5)^2 + \left(\frac{1}{3}-0\right)^2} = 1.05$$

datapoints	cluster k
(0,0)	u_1
(4,0)	u_2
(5,1)	u_2
(1,0)	u_2

5 Fourth iteration

$u_1 = (0, 0)$

$u_2 = \left(5, \frac{1}{3}\right)$

∴ They fall under the same cluster as in the 3 ~~partition~~ iteration

datapoints	cluster k
(0,0)	u_1
(4,0)	u_2
(5,1)	u_2
(6,0)	u_2

6. Fifth Iteration

$u_1 = (0,0)$

$u_2 = (5, \frac{1}{3})$

same as the third iteration.

datapoints	cluster k
(0,0)	u_1
(4,0)	u_2
(5,1)	u_2
(6,0)	u_2

Question 2

1 (1,1)
2 (1,2)
3 (1,5)
4 (4,1)
5 (4,2)
6. (4,5)

Initial points selected are (1,5) (4,2)

Part 1 finding

1. Corresponding quantization errors

For (1,1)

$d_1 = \sqrt{(1-1)^2 + (1-0)^2} = 1$

$d_2 = \sqrt{(4-1)^2 + (2-1)^2} = \sqrt{9+1} = \sqrt{10} = 3.162$

For (1,2)

$d_1 = \sqrt{(1-1)^2 + (2-0)^2} = 2$

$d_2 = \sqrt{(4-1)^2 + (2-2)^2} = 3.$

For (1,0)

$d_1 = 0$

$d_2 = \sqrt{(4-1)^2 + (2-0)^2} = \sqrt{9+4} = \sqrt{13} = 3.61$

For (4,1)

$d_1 = \sqrt{(4-1)^2 + (1-0)^2} = \sqrt{9+1} = \sqrt{10} = 3.162$

$d_2 = \sqrt{(4-4)^2 + (1-2)^2} = 1$

For (4,2)

$d_1 = \sqrt{(4-1)^2 + (2-0)^2} = \sqrt{13} = 3.61$

$d_2 = 0.$

For (4,0)

$d_1 = \sqrt{(4-1)^2 + (0-0)^2} = 3$

$d_2 = \sqrt{(4-4)^2 + (2-0)^2} = 2$

		Cluster
1	(1,1)	u_1
2	(1,2)	u_1
3	(1,0)	u_1
4	(4,1)	u_2
5	(4,2)	u_2
6	(4,0)	u_2

Recalculate centroid.

$$u_1 = \left(\frac{1+1+1}{3}, \frac{1+2+0}{3}\right) = (1,1)$$

$$u_2 = \left(\frac{4+4+4}{3}, \frac{1+2+0}{3}\right) = (4,1)$$

Reassigning

Point 1 (1,1) = u_1
Point 2 (1,2) = u_1
Point 3 (1,0) = u_1
(4,1) = u_2
(4,2) = u_2
(4,0) = u_2

Quantization error.

For (1,1) $\sqrt{(1-1)^2 + (1-1)^2} = 0$

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

For (1,0) = $\sqrt{(1-1)^2 + (1-0)^2} = 1$

For (4,1)

$\sqrt{(4-4)^2+(1-1)^2} = 0$

For (4,2)

$\sqrt{(4-4)^2+(2-1)^2} = 1$

For (4,0)

$\sqrt{(4-4)^2+(2-4)^2} = 4$

Quantization error $= 1+1+0+1+1+1+0 = 4$

clusters

$c(1)=1 \quad c(2)=1 \quad c(3)=1 \quad c(4)=1 \quad c(5)=2$

$c(6)=2.$

2. If $u_1 = (4,1) \quad u_2 = (4,0)$

For (1,1)

$d_1 = \sqrt{(4-1)^2+(1-1)^2} = \sqrt{9} = 3$

$d_2 = \sqrt{(4-1)^2+(1-0)^2} = \sqrt{10} = 3.162$

For (1,2)

$d_1 = \sqrt{(4-1)^2+(1-2)^2} = \sqrt{9+1} = \sqrt{10} = 3.162$

$d_2 = \sqrt{(4-1)^2+(0-2)^2} = \sqrt{9+4} = \sqrt{13} = 3.61$

For (1,0)

$d_1 = \sqrt{(1-4)^2+(1-1)^2} = \sqrt{9+0} = \sqrt{10} = 3.16$

$d_2 = \sqrt{(1-4)^2+(1-0)^2} = \sqrt{9} = 3$

For (4,1)

$d_1 = \sqrt{(4-4)^2 + (1-1)^2} = 0$

$d_2 = \sqrt{(4-4)^2 + (1-0)^2} = 1$

For (4,2)

$d_1 = \sqrt{(4-4)^2 + (2-1)^2} = 1$

$d_2 = \sqrt{(4-4)^2 + (2-0)^2} = 2$

For (4,0)

$d_1 = \sqrt{(4-4)^2 + (0-1)^2} = 1$

$d_2 = \sqrt{(4-4)^2 + (0-0)^2} = 0$

	cluster
(1,1) =	u_1
(1,2) =	u_1
(1,0) =	u_2
(4,1) =	u_1
(4,2) =	u_1
(4,0) =	u_2

Computing new centroid

$u_1 = \left(\frac{1+1+4+4}{4}, \frac{1+2+1+2}{4}\right) = (2.5, 1.5)$

$u_2 = \left(\frac{1+4}{2}, \frac{0+0}{2}\right) = \left(\frac{5}{2}, 0\right) = (2.5, 0)$

7

Rearranging

For (1,1)

$d_1 = \sqrt{(2.5-1)^2 + (1.5-1)^2} = \sqrt{2.5} = 1.58$

$d_2 = \sqrt{(2.5-1)^2 + (1-0)^2} = \sqrt{3.25} = 1.80$

For (1,2)

$d_1 = \sqrt{(2.5-1)^2 + (1.5-2)^2} = 1.58$

$d_2 = \sqrt{(2.5-1)^2 + (0-2)^2} = \sqrt{(1.5)^2 + 4} = 2.5$

For (1,0)

$d_1 = \sqrt{(2.5-1)^2 + (1.5-0)^2} = 2.121$

$d_2 = \sqrt{(2.5-1)^2 + (0-0)^2} = 1.5$

For (4,1)

$d_1 = \sqrt{(4-2.5)^2 + (1.5-1)^2} = 1.58$

$d_2 = \sqrt{(4-2.5)^2 + (1-0)^2} = 1.802$

For (4,2)

$d_1 = \sqrt{(4-2.5)^2 + (2-1.5)^2} = 1.58$

$d_2 = \sqrt{(4-2.5)^2 + (2-0)^2} = \sqrt{6.25} = 2.5$

For (4,0)

$d_1 = \sqrt{(4-2.5)^2 + (1.5-0)^2} = 2.121$

$d_2 = \sqrt{(4-2.5)^2 + (0-0)^2} = 1.5$

c(1) = 1 c(2) = 1 c(3) = 2 c(4) = 1 c(5) = 1
c(6) = 2

Quantization error:

$E = (1.58)^2 + (1.58)^2 + (1.5)^2 + (1.58)^2 + (1.58)^2 + (1.5)^2$

$E = 14.5$

Part-2

1. given
1 (1,1)
2 (1,2)
3 (1,0)
4 (4,1)
5 (4,2)
6 (4,0)

Point 1 (1,1) is selected.

Calculate distances

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

For (1,0)
$\sqrt{(1-1)^2 + (1-0)^2} = 1$

for (4,1) $= \sqrt{(4-1)^2 + (1-1)^2} = 3$

For (4,2) $= \sqrt{(4-1)^2 + (2-1)^2} = \sqrt{9+1} = \sqrt{10} = 3.16$

For (4,0) $= \sqrt{(4-1)^2 + (0-1)^2} = \sqrt{9+1} = \sqrt{10} = 3.16$

$Z = 0+1+1+3+3.16+3.16$

1.

1	$d = 6$
2	$d = 1$
3	$d = 1$
4	$d = 3$
5	$d = 3.16$
6	$d = 3.16.$

2. $Z = 0+1+1+3+3.16+3.16 = 11.32.$

Calculating probability:

For $(1,1) = \frac{0}{11.32} = 0$

For $(1,2) = \frac{1}{11.32} = 0.088$

For $(1,0) = \frac{1}{11.32} = 0.088$

For $(4,1) = \frac{3}{11.32} = 0.265$

For $(4,2) = \frac{3.16}{11.32} = 0.279$

For $(4,0) = \frac{3.16}{11.32} = 0.279$

1	$P = 0$
2	$P = 0.088$
3	$P = 0.088$
4	$P = 0.265$
5	$P = 0.279 \approx 0.28$
6	$P = 0.279 \approx 0.28$

3. The second point can be points 5 or points 6 based on the largest probability

Choosing point 6 as second point (4,0)

$u_1 = (1,1)$ $u_2 = (4,0)$

Calculating distances.

For (1,1)

$d_1 = 0$

$d_2 = \sqrt{(4-1)^2 + (1-0)^2} = \sqrt{9+1} = \sqrt{10} = 3.16$

For (1,2)

$d_1 = 1$

$d_2 = \sqrt{(4-1)^2 + (2-0)^2} = \sqrt{9+4} = \sqrt{13} \approx 3.61$

For (1,0)

$d_1 = 1$

$d_2 = \sqrt{(4-1)^2 + (0-0)^2} = \sqrt{9+0} = 3$

For (4,1)

$d_1 = 3$

$d_2 = \sqrt{(4-4)^2 + (1-0)^2} = 1$

For (4,2)

$d_1 = 3.16$.

$$d_2 = \sqrt{(4-4)^2 + (2-0)^2} = \sqrt{4} = 2$$

For (4,0)

$d_1 = 3.16$

$d_2 = 0$

$c(1) = 1 \quad c(2) = 1 \quad c(3) = 1 \quad c(4) = 2$

$c(5) = 2. \quad c(6) = 2$

~~B~~ =

~~Calcul~~ Calculating new centroid

$$u_1 = \left(\frac{1+1+1}{3}, \frac{1+2+0}{3}\right) = (1,1)$$

$$u_2 = \left(\frac{4+4+4}{3}, \frac{1+2+0}{3}\right) = (4,1)$$

Quantization error.

$E = (0)^2 + (1)^2 + (1)^2 + 9 + 0 + 4 = 15$

4. If the algorithm selects the point with the second largest probability points (4,2) then yes, it could change the initial clustering. ~~and both the points assigned~~ and points could potentially be assigned to different clusters.

5. It can cause different clustering after convergence & a different quantization error as well. The initial cluster centers would be point1 (1,1) & point4 (4,1)

6. The probability that the algorithm chooses
point 4 is 0.265
point 5 is 0.279
point 6 is 0.279

Question3

given, department = systems
status = senior
age = 21-30.
Count(salary = low) = 3 Count(salary = high) = 2
P(department = systems | salary = low) = 0
P(department = systems | salary = medium) = $\frac{2}{6} = \frac{1}{3}$
P(department = systems | salary = high) = $\frac{2}{2} = 1$

P(status = senior | salary = low) = 0.
P(status = senior | salary = medium) = $\frac{3}{6} = \frac{1}{2}$
P(status = senior | salary = high) = $\frac{2}{2} = 1$

P(age = 21-30 | salary = low) = $\frac{2}{3}$
P(age = 21-30 | salary = medium) = $\frac{2}{6}$
P(age = 21-30 | salary = high) = 0.

Prior probability of salary = low = $\frac{3}{11}$
Prior probability of salary = medium = $\frac{6}{11}$
Prior probability of salary = high = $\frac{2}{11}$

For salary = low

= Prior prob × P(department = systems | salary = low) × P(de status = senior | salary = low) × P(age = 21-30 | salary = low)

$= 0.272 \times 0 \times \frac{1}{6} \times 0 \times \frac{2}{3} = 0.$

For salary = medium:

= Prior probability × P(dept = systems, status = senior, age = 21-30 | salary = medium)

$= \frac{6}{11} \times \frac{2}{6} \times \frac{3}{6} \times \frac{2}{6} = \frac{12}{396} = 0.03030.$

For salary = high.

= Prior prob × P(dept = systems, status = senior age = 21-30 | salary = high)

$= \frac{2}{11} \times 1 \times 1 \times 0 = 0.$

Therefore, the naive bayes classification for the salary of the given instance is salary = medium //