ML Assignment - 3

Q.1

Ans. Given: data points = $(-1,3)$, $(-3,1)$ and $(-2,-1)$

The new centroid of the points

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

Centroid $= (-2, 1)$ Ans

Q.2

Ans. Given: data points: $A_1(2,10)$, $A_2(2,5)$, $A_3(8,4)$
$A_4(5,8)$, $A_5(7,5)$, $A_6(6,4)$, $A_7(1,2)$
$A_8(4,9)$ ~~Iterga~~ Iteration 1

(i) No. of clusters $K = 3$

(ii) Initializing centroids $k_1 = (2,5)$
$k_2 = (4,9)$
$k_3 = (7,5)$

(iii) Assigning the all data points into a cluster

$A_1 \rightarrow D_1 = \{(2,10), (2,5)\}$
$= \sqrt{(2-2)^2 + (10-5)^2} = 5$

$D_2 = \{(2,10), (4,9)\}$
$= \sqrt{(2-4)^2 + (10-9)^2} = 2.236$

$D_3 = \{(2,10), (7,5)\}$
$= \sqrt{(2-7)^2 + (10-5)^2} = 7.07$

So point A_1 is nearer to k_2, so we will assign it into k_2 cluster.

Similarly for all data points

$A_2 \rightarrow D_1 = 0$, $D_2 = 4.47$, $D_3 = 5$

A_2 in cluster k_1

$A_3 \rightarrow D_1 = 6.08$, $D_2 = 6.4$, $D_3 = 1.414$

A_3 in cluster k_3

$A_4 \longrightarrow D_1 = 4.24$, $D_2 = 1.414$, $D_3 = 3.61$

A_4 in cluster K_2

$A_5 \longrightarrow D_1 = 5$, $D_2 = 5$, $D_3 = 0$

A_5 in cluster K_3

$A_6 \longrightarrow D_1 = 4.123$, $D_2 = 5.38$, $D_3 = 2.24$

A_6 in cluster K_3

$A_7 \longrightarrow D_1 = 3.16$, $D_2 = 7.62$, $D_3 = 6.71$

A_7 in cluster K_1

$A_8 \longrightarrow D_1 = 4.47$, $D_2 = 0$, $D_3 = 5$

A_8 in cluster K_2

So K_1 cluster $= (2,5), (1,2)$

Centroid $= \left(\frac{2+1}{2}, \frac{5+2}{2}\right) = (1.5, 3.5)$

K_2 cluster $= (2,10), (5,8), (4,9)$

Centroid of $K_2 = \left(\frac{2+5+4}{3}, \frac{10+8+9}{3}\right)$

Centroid of $K_2 = (3.67, 9)$

K_3 cluster $= (8,4), (7,5), (6,4)$

Centroid of $K_3 = \left(\frac{8+7+6}{3}, \frac{4+5+4}{3}\right)$

Centroid of $K_3 = (7, 4.33)$

Iteration 2

New Cluster Centroids $K_1 = (1.5, 3.5)$

$K_2 = (3.67, 9)$

$K_3 = (7, 4.33)$

$A_1 \rightarrow D_1 = 6.519$, $D_2 = 1.95$, $D_3 = 7.56$

A_1 in K_2 cluster same as before

$A_2 \rightarrow D_1 = 1.58$, $D_2 = 4.33$, $D_3 = 5.04$

A_2 in K_1 cluster

$A_3 \rightarrow D_1 = 6.52$, $D_2 = 6.61$, $D_3 = 1.05$

A_3 in K_3 cluster

$A_4 \rightarrow D_1 = 5.7$, $D_2 = 1.66$, $D_3 = 4.18$

A_4 in K_2 cluster

$A_5 \rightarrow D_1 = 5.7$, $D_2 = 5.2$, $D_3 = 0.67$

A_5 in K_3 cluster

$A_6 \rightarrow D_1 = 4.53$, $D_2 = 5.52$, $D_3 = 1.05$, A_6 in K_3

$A_7 \rightarrow D_1 = 1.58$, $D_2 = 7.49$, $D_3 = 6.44$

A_7 in cluster K_1

$A_8 \rightarrow D_1 = 6.04$, $D_2 = 0.33$, $D_3 = 5.55$

A_8 in cluster K_2

so final clusters :→ $K_1 = (2, 5), (1, 2)$

$K_2 = (2, 10), (5, 8), (4, 9)$

$K_3 = (8, 4), (7, 5), (6, 4)$

Ans.

Q. (3)

Ans. All three Conditions Can act as Possible termination Conditions in k-means.

~~But for Case (2) except for the Cases with a bad local minimum~~

But for Condition (2) except for the Cases with bad local minimum, this Produces a good clustering, but runtime may be unacceptably long.

Q. (4)

Ans. Yes, k-means is sensitive to outliers because it is Considering mean value of cluster and if a outlier is added to cluster then it will change the mean drastically.

Q. (5)

Ans. (a) support Count of {Milk, Diaper, Bread}

$= 2$ Ans.

(b) Support of {Milk, Diaper, Bread}

$$= \frac{2}{5} = 0.4$$

(c) (i) {Milk, Diaper} $\longrightarrow$ {Beer}

$$\text{support } s = \frac{2}{5} = 0.4$$

$$\text{Confidence } c = \frac{s(\text{Milk, Diaper, Beer})}{s(\text{Milk, Diaper})} = \frac{2}{3}$$

$$c = 0.67$$

(ii) {Milk, beer} $\longrightarrow$ {Diaper}

$s = 0.4$ (same as before)

$$\text{Confidence } c = \frac{s(\text{Milk, Diaper, Beer})}{s(\text{Milk, beer})} = \frac{2}{2} = 1.0$$

(iii) {Diaper, beer} $\longrightarrow$ {Milk}

$s = 0.4$

$$c = \frac{s(\text{Milk, Diaper, Beer})}{s(\text{Diaper, beer})} = \frac{2}{3} = 0.67$$

(iv) {Beer} $\longrightarrow$ {Milk, Diaper}

$$S = 0.4, \quad c = \frac{S(\text{Milk, Beer, Diaper})}{S(\text{Beer})} = \frac{2}{3} = 0.67$$

(v) {Diaper} $\longrightarrow$ {Milk, beer}

$$S = 0.4, \quad c = \frac{S(\text{Milk, Beer, Diaper})}{S(\text{Diaper})} = \frac{2}{4} = 0.5$$

(vi) {Milk} $\longrightarrow$ {Diaper, Beer}

$$S = 0.4, \quad c = \frac{S(\text{Milk, Beer, Diaper})}{S(\text{Milk})}$$

$$c = \frac{2}{4} = 0.5$$ Ans.

Q.6

Ans. (a) False, 2^d Possible itemsets

(b) True

(c) False

(d) True

Q.7

Ans. a) support of item {e} $= \frac{8}{10} = 0.8$

support of itemset {b,d} $= \frac{2}{10} = 0.2$

support of itemset {b,d,e} $= \frac{2}{10} = 0.2$

b) Confidence of itemsets {b,d} $\rightarrow$ {e}

$$c = \frac{S(b,d,e)}{S(b,d)} = \frac{0.2}{0.2} = 1$$ Ans.

Confidence of itemsets {e} $\rightarrow$ {b,d}

$$c = \frac{S(b,d,e)}{S(e)} = \frac{0.2}{0.8} = 0.25$$ Ans.

Q. 8

Ans. $\{\text{all frequent Patterns}\} \geq \{\text{closed frequent Patterns}\} \geq \{\text{max. frequent Patterns}\}$

Q. 9

Ans. Given: support of itemset $\{a, b, c\} = 10$

means that support of itemset $\{a, b\} \geq$ support of itemset $\{a, b, c\}$

so support of itemset $\{a, b\}$ that is Possible is ≥ 10

$S\{a, b\} = 10$ or 11 or 12 Ans.