CSE 3024

Web Mining

DIGITAL ASSIGNMENT - 1

NAME: Vibhu Kumar Singh
REG. NO: 19BCE0215
TEACHER: Mr. Hiteshwar Kumar Azad

1. To divide the given dataset into two clusters, use the K means clustering algorithm. Assume that the initial centroids are (40, 71) and (36, 68). Utilize the Euclidean metric. Iterate two times.

City	Temperature (°C)	Humidity (%)
Vellore	39	67
Chennai	40	71
Patna	33	54
Banaras	32	53
Mumbai	35	66
Vijayawada	36	68
Bareli	32.5	60
Bhopal	31	51

Ans 1.

Q1)

City	Temperature	Humidity
Vellore	39	67
Chennai	40	71
Patna	33	54
Banaras	32	53
Mumbai	35	66
Vijaywada	36	68
Bareli	32.5	60
Bhopal	31	51

using Euclidean metric

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

Given centroid

$$C_1 (40, 71)$$
$$C_2 (36, 68)$$

City wise distance from centroids

* Vellore

P(39, 67)
$C_1(40, 71)$
$C_2(36, 68)$

$$Dis (C_1) = \sqrt{(40 - 39)^2 + (71 - 67)^2}$$
$$= \sqrt{1 + 4^2}$$
$$= \sqrt{17}$$

$$Dis (C_2) = \sqrt{(36 - 39)^2 + (68 - 67)^2}$$
$$= \sqrt{3^2 + 1^2}$$
$$= \sqrt{10}$$

* Chennai

P(40, 71)
$C_1(40, 71)$
$C_2(36, 68)$

$$Dis (C_1) = \sqrt{(40 - 40)^2 + (71 - 71)^2}$$
$$= \sqrt{0 + 0} = 0.$$

$$Dis (C_2) = \sqrt{(36 - 40)^2 + (68 - 71)^2}$$
$$= 5$$

⋆ Patna

P(33,54)

$C_1(40,71)$

$C_2(36,68)$

$Dis(C_1) = \sqrt{(40-32)^2 + (71-54)^2}$

$= \sqrt{7^2 + 17^2}$

$= \sqrt{338}$

$Dis(C_2) = \sqrt{(36-33)^2 + (54-60)^2}$

$= \sqrt{9+196}$

$= \sqrt{205}$

⋆ Banaras

P(32, 53)

$C_1(40,71)$

$C_2(36,68)$

$Dis(C_1) = \sqrt{(40-32)^2 + (71-53)^2}$

$= \sqrt{388}$

$Dis(C_2) = \sqrt{(36-32)^2 + (68-53)^2}$

$= \sqrt{241}$

⋆ Mumbai

P(35,66)

$C_1(40,71)$

$C_2(36,68)$

$Dis(C_1) = \sqrt{(40-35)^2 + (71-68)^2}$

$= \sqrt{50}$

$Dis(C_2) = \sqrt{g^2 + 2^2}$

$= \sqrt{5}$

⋆ Vijaywada

P(36,68)

$C_1(40,71)$

$C_2(36,68)$

$Dis(C_1) = \sqrt{(40-36)^2 + (71-68)^2}$

$= 5$

$Dis(C_2) = \sqrt{0^2 + 0^2}$

$= 0.$

⋆ Bareli°

P(32.5, 60)

$C_1(40,71)$

$C_2(36,68)$

$Dis(C_1) = \sqrt{(40-32.5)^2 + (71-60)^2}$

$= \sqrt{177.25}$

$Dis(C_2) = \sqrt{(36-32.5)^2 + (68-60)^2}$

$= \sqrt{76.25}$

⋆ Bhopal

P(31,51)

$C_1(40,71)$

$C_2(36,68)$

$Dis(C_1) = \sqrt{(40-31)^2 + (71-51)^2}$

$= \sqrt{481}$

$Dis(C_2) = \sqrt{(36-31)^2 + (68-51)^2}$

$= \sqrt{314}$

City	Dis for C_1	Dis for C_2	Cluster
Vellore	$\sqrt{17}$	$\sqrt{10}$	C_2
Chennai	0	$\sqrt{25}$	C_1
Patna	$\sqrt{338}$	$\sqrt{205}$	C_2
Banaras	$\sqrt{388}$	$\sqrt{241}$	C_2
Mumbai	$\sqrt{50}$	$\sqrt{5}$	C_2
Vijaywada	$\sqrt{25}$	0	C_2
Bareli	$\sqrt{177.25}$	$\sqrt{76.25}$	C_2
Bhopal	$\sqrt{481}$	$\sqrt{314}$	C_2

New centroids

for $C_1 = \left(\dfrac{40}{1}, \dfrac{71}{1} \right)$

for $C_2 = \dfrac{39 + 33 + 32 + 35 + 36 + 32.5 + 31}{7} = \dfrac{238.5}{7} = 34.07$

$\dfrac{67 + 54 + 53 + 66 + 68 + 60 + 51}{7} = \dfrac{419}{7} = 59.85$

After iteration 1

Centroid $C_1 = \{(40, 71)\}$

Cluster $= \{(40, 71)\}$

Centroid $C_2 = (34.07, 59.85)$

Cluster $= \{(39, 67), (33, 54), (32, 53)(35, 66),$
$(36, 68), (32.5, 60), (31, 51)\}$

Iteration 2

✱ Vellore

$P(34, 67)$

$C_1(40, 71)$

$C_2(34.07, 59.85)$

$Dis(C_1) = \sqrt{(40-39)^2 + (71-67)^2}$

$= \sqrt{17}$

$Dis(C_2) = \sqrt{(34.07-34)^2 + (59.85-67)^2}$

$= \sqrt{75.42}$

✱ Chennai

$P(40, 71)$

$C_1(40, 71)$

$C_2(34.07, 59.85)$

$Dis(C_1) = \sqrt{(40-40)^2 + (71-71)^2}$

$= 0.$

$Dis(C_2) = \sqrt{(34.07-40)^2 + (59.85-71)^2}$

$= \sqrt{159.48}$

✱ Petua

$P(33, 54)$

$C_1(40, 71)$

$C_2(34.07, 54.85)$

$Dis(C_1) = \sqrt{(40-33)^2 + (71-54)^2}$

$= \sqrt{338}$

$Dis(C_2) = \sqrt{(34.07-33)^2 + (59.85^2 - 54)}$

$= \sqrt{35.36}$

✱ Banaras

$P(32, 53)$

$C_1(40, 71)$

$C_2(34.07, 59.85)$

$Dis(C_1) = \sqrt{(40-32)^2 + (71-53)^2}$

$= \sqrt{388}$

$Dis(C_2) = \sqrt{(34.07-32)^2 + (59.35-53)^2}$

$= \sqrt{51.2}$

✱ Mumbai

$P(35, 66)$

$C_1(40, 71)$

$C_2(34.07, 59.85)$

$Dis(C_1) = \sqrt{(40-25)^2 + (71-66)^2}$

$= \sqrt{50}$

$Dis(C_2) = \sqrt{(34.07-35)^2 + (59.85-66)^2}$

$= \sqrt{38.68}$

* Vijaywada

P(36,68)

C₁(40,71)

C₂(34.07, 59.85)

$$Dis(C_1) = \sqrt{(40-36)^2 + (71-68)^2}$$
$$= \sqrt{25}$$

$$Dis(C_2) = \sqrt{(34.07-36)^2 + (59.85-68)^2}$$
$$= \sqrt{70.14}$$

* Bareli

P(32.5, 60)

C₁(40, 71)

C₂(34.07, 59.85)

$$Dis(C_1) = \sqrt{(40-32.5)^2 + (71-60)^2}$$
$$= \sqrt{177.25}$$

$$Dis(C_2) = \sqrt{(34.17-32.5)^2 + (59.85-60)^2}$$
$$= \sqrt{2.48}$$

* Bhopal

P(31,51)

C₁(40,71)

C₂(34.07, 59.85)

$$Dis(C_1) = \sqrt{(40-31)^2 + (71-51)^2}$$
$$= \sqrt{481}$$

$$Dis(C_2) = \sqrt{(34.07-31)^2 + (59.85-51)^2}$$
$$= \sqrt{87.74}$$

City	Dis from C₁	Dis from C₂	Cluster
Vellore	$\sqrt{17}$	$\sqrt{75.42}$	C₁
Chennai	0	$\sqrt{159.48}$	C₁
Patna	$\sqrt{338}$	$\sqrt{35.36}$	C₂
Banares	$\sqrt{388}$	$\sqrt{51.2}$	C₂
Mumbai	$\sqrt{56}$	$\sqrt{38.68}$	C₂
Vijaywada	$\sqrt{25}$	$\sqrt{70.14}$	C₁
Bareli	$\sqrt{177.25}$	$\sqrt{2.48}$	C₂
Bhopal	$\sqrt{481}$	$\sqrt{87.74}$	C₂

New Cluster with cluster head

Cluster 1 = { (39,67), (40,71), (36, 68) }

$$C_1 = \frac{39+40+36}{3} = \frac{115}{3} = 38.33$$

$$\frac{67+71+68}{3} = \frac{206}{3} = 68.67$$

$$\boxed{C_1 = (38.33, 68.67)}$$

Cluster 2 = { (33,54), (32,53), (35,66)
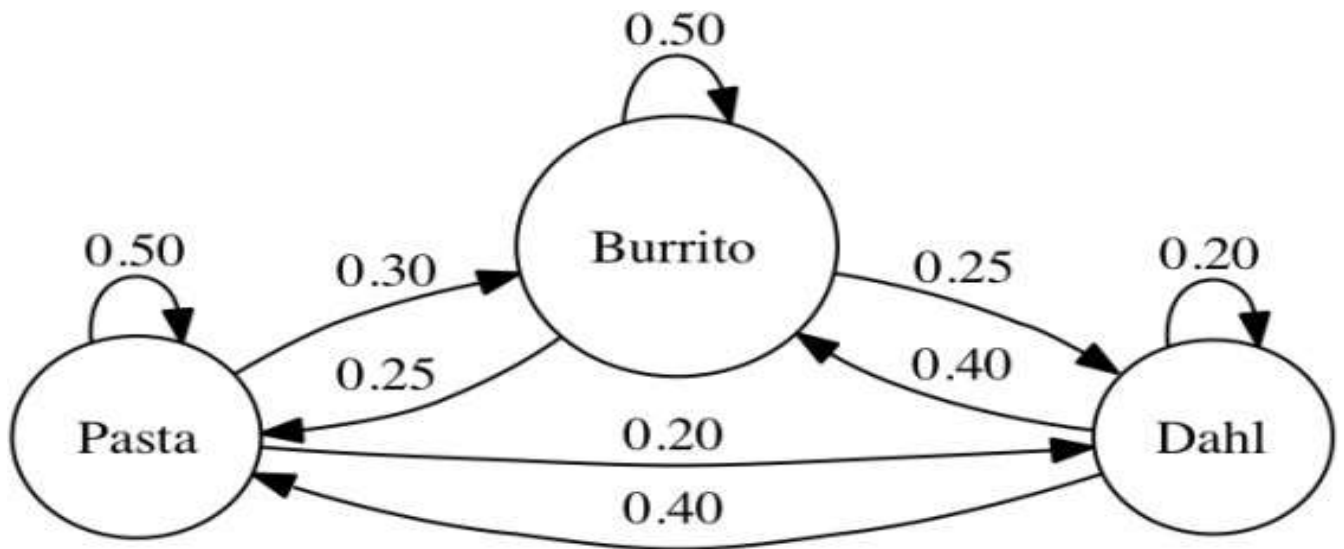(32.5, 60), (31,51) }

$$C_2 = \frac{33+32+35+32.5+31}{5} = \frac{163.5}{5} = 32.9$$

$$\frac{54+53+66+60+51}{5} = \frac{284}{5} = 56.8$$
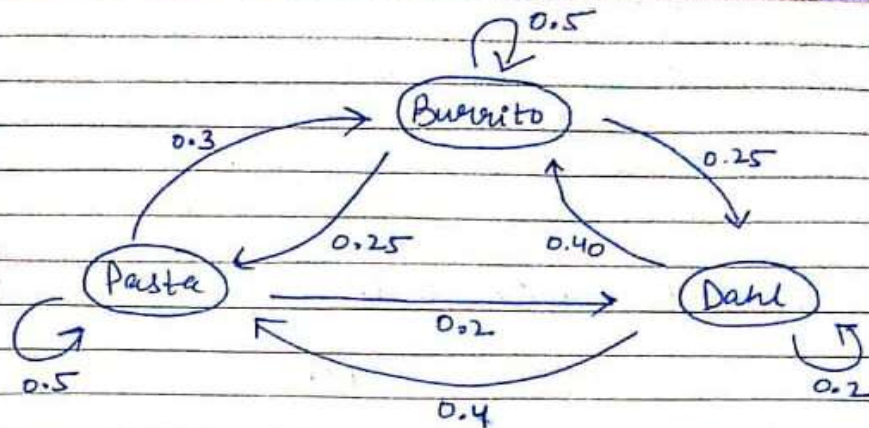
$$\boxed{C_2 = (32.9, 56.8)}$$

Ans

2. Draw a State-transition probability matrix using the given Markov Chain to show the transition probabilities of each state. If anyone eats Dahl today, what are the chances that pasta will be served tomorrow, and burritos will be served the next day?

Ans 2.

(P.T.O)

Q2)

The state transition probability matrix for the above markov chain :-

	Burrito	Pasta	Dahl	
Burrito	0.5	0.25	0.25	= 1
Pasta	0.3	0.5	0.2	= 1
Dahl	0.4	0.4	0.2	= 1

Symbols

Burrito $\Rightarrow$ B

Pasta $\Rightarrow$ P

Dahl $\Rightarrow$ D

$P(B/B) = 0.5$ $P(D/P) = 0.2$

$P(P/B) = 0.25$ $P(B/D) = 0.4$

$P(D/B) = 0.25$ $P(P/D) = 0.4$

$P(P/P) = 0.5$ $P(D/D) = 0.2$

Query $\Rightarrow$ Today = Dahl,
             Tommorrow = Pasta,
       Day after Tomorrow = Burritoes

Probability $= P(B/P, D)$

$\qquad = P(B/P) \times P(P/D)$

$\qquad = 0.3 \times 0.4$

$\qquad = \boxed{0.12}$

$\qquad\qquad\qquad\underline{Ans}$