11415V) 834V/ %

12. Deline Clustering a set of clustering is the Process of grouping instances into clusters. So that instances within a cluster have high similarity

13. Little the applications for clustering The applications of clustering and a. Human genetic dustering

b. Medical genetic Clustering

c. Mankel he- search d. Pattern ne-cognition

10,3/ 4

14. Write the requirements of Clustering

The recovarinements of clustering are

a. Dealing with large data sets Containing different types of attributes.

b. Final the clusters with arbitrary

c. Ability to ded with noisy data in data streaming environment.

d. Handling with high dimensional data sets.

H Clustening HA MAGAG TOWN SILLS Clustering > Partitioning methods -> Hie manchical methods Density - based methods olati files cost 2 morelans in get (interior) Grid based methods and it is instant to a planton is in one

15. Describe different clustering methods

Ans

110 9109

Whene k L=N.

of lessence and declared off post offered of N-> means numbers of instances.

It linds mutually exclusive clusters of spo spherical shape using transfirmal distance measures. Effective son clustering high-dimensional big data.

Example: K-means Clustering, K-medoids

b. Hieroandical m methods: It creates a hieroanchical It has two categories decomposition of 1 instances.

143/14

having all the N instances & then split into Smaller clusters in each successive iteration until eventually each instances is in one clusters. on termination condition holds.

ii. Bottom up: Fach instance forom a separate clusters. Then the clusters are menged to one another. It continues until all the clusters are menged into a single clusters.

one tenmination conditions holds.

Example: Age Agres, BIRCH.

Example: 10- ments clustering, K. medoids

C. Density based method: clustering instances based on the distance between instances. It linds an bi-inamily shaped clusters. It can cluster instances as dense regions in data space.

d. Graid based method: The anid historial user uses a multi nesolution good data structure. It's independent of numbers of instances.

16. Haite the route of similarity measure

Ans.

The similarity roule solute solute

a. Each clusters must contain at least one instance.

b. Fach instance must belong to exactly one

17. Write the Property of C-mego

High degree of similarity among instances in clusters is obtained.

High alegace of dissimilarity among instances in different Clusters is achieved simul-laneously

18. Hitie K-mens clustering algorithm. Input: (X = 1x1, xa, to bxar) liandidas 11 A set of unlabelled instance. Output: A set of k clusters.

Method: Method:

775-71-19

The independent of number of 1. An bi-transity choose k numbers of instances, d xk,, xka, ..., xkr & x as the inflial K clusters centers de glissiano de

Ital to makes there asked that

3. Re-assign each X: EX-K, someon value Xm Ek.

9. Update the lack means that is calculate the mean value of the instances for each grand at elustes, plinoliniais to magala doil 5. Until no change

19. Write the aboutback of k-means Clustering 104

. The drawbacks are

the of

0000 a Not guranteed to converge to the global optimum

b. Can not be used in data with nominal Jealunes.

c. Not suitable for dis-covering clusters.

do Fixed number of clusters can make it diff difficult to Predict what k should be.

e. The number of nkl). n -> Total number of instances. K+ number of clusters. L + Numbers 1.0 = 2(81-1-1) of itenations.

1(1.88-110) = 0 62 = 0 (1.88-89) = 0 35

10:5(8:0- 20) 10 = 5(21-51) 1-

30-5(88-51)1 0 = 5(0.0-0.0) 1. 0 = 5(0.16 0.0) 1

20 - 7(11-1-0) 1 10 = 5(0,0-1,9) 1

1 0 = 5(3 8 - 3 1) 1.

350 - 500 - 3011

E 9 = 1 2 (3-) - CO 10

Table: Height Data

Name	Gender	Height	Output
Kristina	F	1.6 m	Short
Jim	M	2 m	Tall
Maggie	F	1.9 m	Medium
Martha	F	1.88 m	Medium
Stephanie	F	1.7 m	Short
Bob	M	1.85 m	Medium
Kathy	F	1.6 m	Short
Dave	M	1.7 m	Short
Worth	M	2.2 m	Tall
Steven	M	2.1 m	Tall
Debbie	F	1.8 m	Medium
Todd	M	1.95 m	Medium
Kim	F	1.9 m	Medium
Amy	F	1.8 m	Medium
Wynette	F	1.75 m	Medium

Do K means clustering on k=2

-			
FIG.			Pojen
SHEF	let k	shocks are converged to be	a Not guranter
			b. Con not b
Senia)	Height		plat Distance from Cz
1	1.6	1 (1.6-1.6)2 = 0	16 (1.6-2.2) 2 = 0 6
3		1 (2-1-6)2 = 0.4	
3	1.9	7(1.9-1.6)2 = 0.3	J(1.9-2.2)2 = 0.3
9	1.88 mine	1 (1.88-1.6)2 = 0.28	1(1.88-2.2)2:= 0.32
5190	mulg-7-	1(1.7-1.6)2 = 0.1	1(1.7-2.5)2 = 0.5
6	1.85	1(1.85-1.6)2 = 0.25	, (1.85 - 2.2) 2 = 0.35
7	1.6	1(1.6-1.6)2 = 0	
8	1.7	1 (1-7-1-6)2 = 0.1	1(1.6 - 5.5) 5 = 0.6
9	2.2	1 (2.2-1.6)2 = 0.6	1(1.4-2.8)2 = 0.5
10	٩٠١	1 (8.1-1.4)2 = 0.5	7 (2.2-2.5)2 = 0
11	1.8	1118-1112	1 (2.1-2.2)2 = 0.1
12	1.95	1(1.8-1.6)2 = 0.2	7(1.8-2.2)2 = 1.4
13	1.9	1(1.95-1.6)2 = 0.35	1 (1.95 - 2.212
14	1.8	1 (1.9-1.6)2 = 0.3	1 - 2 - 9 1 2
15	1-75	1 (1-75 - 1.6)2 = 0.15	1 . 0 - 5 . 3 . 2
1	17		1(1.42-5.5)5 = 0.42

Drugger 1 672 19 water persolated to the Harris House

C1 = 1 1, 3, 4, 5, 6, 7, 8, 11, 13, 14, 15}

Ca = 1 a, 9, 10, 125 \$10 - 5(00-2-881) - 10- 5(0F1 381) -

SIPP:20 = 9(00) 4 - PILL FOR - 9(181-18-1) Finding the means o soll 1 28 11.

C1 = 1.6 + 1.9 + 1.88 + 1.7 + 1.85 + 1.6 + 1.7 + 1.8 + 1.9 + 1.8 + 1.7 +

10 - 1009 - 49148 (100 - (11-1-2 4) 1

(60-1-1-1)2 0.33 dilest - 6 cc)2

= 1.77

2.06 = 2.06

1 (1-28 1-38) = 005 9 (1-12 6.00) 5 = 0

21 91 81 11 01 8 8 10 11 13 19 15

0.1

3	/1			- 1
/	Servial	Heighl.	Distance Ison e/	Distance Isom Ca
	1	1.6 (8) 11	J(1.6+1.77)2=0.17	
	3	2	1(2-1-77)2=0.23	1(2-2.06)2 = 0.06
	3	1.9	109-1-77/2=0.13	1(1.9-2.06)2 = 0-16
	4	1.88	1(1.88-1-77)2 = 0.11	1 (1.88 - 5.00) = 0.18
4	5	1-7	1(1.7-1.77)2= 0.07	J(1.7 - 2.06)2 = 0.36 %
	6	1.85	J(1.85-1.70)2= 0.08	1 (1.85 - 2.06) 2 = 0.21
	₹	1.6	1(1.6-1.77)2=0.17	
-	8	ा ने शिक्षेत्र के । ने देश	J(1.7 - 1.77)2 = 0.07	J(1.6 - 2.06)2 = 0.96
	9	2.2		1(1-7-2.06)2 = 0.36
	16	2.1	1 (2.2-1.77)2= 0.43	4 (8.5-5.00) = 0.19
	11	1.8	V[1.8-1.71]2=0.33	1(2.1-2.06)2 = 0.04
1	12	1.95	10.05 - 1.7712 - 0.10	1(1.8 - 2.06)2 = 0.26
	13	1-9	V(19-1:7712-0.18	1(1.95-2.06)2=0.11
	14	1.8	1(1.8-1.77)2 = 0.03	11.32-2.06)2 = 0.16
	15	1.75	11:75-1:7712-002	1(1.8 - 2.06) 2 = 0.26
1	-		1. 10 , 10 = 0.05	1(1.75 - 2.06)2 = 0.3

C1=11, 3, 4, 5, 6, 7, 8, 10, 11, 13, 19, 15}
C2={2, 9, 12.}

Finding Mean $C_{1}' = \frac{1.6 + 1.9 + 1.88 + 1.7 + 1.85 + 1.6 + 1.7 + 2.1 + 1.8 + 1.9 + 1.8 + 1.7}{12}$ $C_{2}' = \frac{2 + 2.2 + 1.95}{3}$

50, e'_ = 1.79 e'_ = 2.05

Sero ial	Height	Distance Inorn e,1	Distance Inon (2'
1	1.6	J(1.6-1.79) 2 = 0.19	1(16-2.05)2 = 0.45
2	2	J(2-1-79)2 = 0-21	1(2-2.05)2= 0.05
3	1.9	J(1.9-1.79)2 = 0.11	11.9-2.05)2= 0.15
4	1.88	1(1.88 - 1.79)2 = 0.09	1(.88 - 2.05)2 = 0.17
5	1.7	J(1.7-1.79)2 = 0.09	1(1.7-2.05)2 = 0.35
6	1.85	1(1.85-1.79)2 = 0.06	1(1.85 - 2.05)2 = 0.2
Ŧ	1.6	V(1.6-1.79)2 = 0.19	1(1.6-2.05)2 = 0.45
8	1.7	V(1.7-1.79)2 = 0.09	1(1-7-2.05)2 = 0.35
9	2.2	V(2.2-1.79)2 = 0.41	1(2.2-2.05)2 = 0.15
10	2.1	1(2.1-1.79)2 = 0.31	1(2.1-2.05)2 = 0.05
11	1.8	1 (1.8 - 1.79)2 = 0.0	J (1.8-2.05)2 = 0.25
12	1.95	1 (1.95 -1.79)2 = 0.16	1(1.95 - 2.05)2 = 0.1
13	1.9	VI.9 - 1.79/2 = 0.11	J(1.9 - 2.05)2 = 0.11
19	1.8	11.8 -1.79)2 = 0.01	1 (1.8 - 2.05)2 = 0.25
15	1.75	1 (1.75 - 1.79)2 =0	.04 -(1.75 - 2.05) 2 = 0

```
C1=11,3,4,5,6,7,8,10,11,12,13,14,15)
```

2115116

e' = 1-79 e' = 1-79

1 pi ora Ca monte 30 moters Heigisell 0.1 1(16-2.05)2 1(16-1-19)2 : 019 1 % 18.0 = s(ex 1-s) L 1(2-8-05)2= 6.1 91 - 5 (00 % - 81) 11.0 -1.70)2 -0.11 600 - 6 (6) 1-88-1) 1-38 1 11.88 - 2.05)2 = (0.0 s(e)-1-51). = 0(30.9 - F-1)} 1-1 200-5(61.1-98-1) - 5(20.9 - 83.1)+ 98-1 31 C10 = 2(0) 1 - 21) 1 1(1.6.0.0)2 = 2(20.5 - 4-1) } 60.0 = c(62.1-21)p 1(20-2-05)2: 110 30 (621-00) 2.2 2 5 (201 - 205) 2 = 18.0: 5(0) 1-105% 1. 5 10.0 - 5(0) 1-3-1) 1 = 8 (20 8 - 8-1) } 8.1 110 - 5101-1- 1911 4(195-205): 1.95 1.0 110 = 2 (6) 1 - 6.16 110 1 2 (30 0 - 2 1) 6.4 100 - 5168 1 - 8.11 4 (118-6-00) = 1000 1 11:15 - 1 79) + 10:0 - 1(ex - excs) = 0.5