DESCRIPTION

Subjects are divided into different classes such as theoretical, logical, programming and technical subjects . For each student find the mean marks he/she scored for theoretical subjects, logical subjects, programming, technical subjects separately .

DATA

theoritical	logical	programming	technical
physics	Maths	Computer programming lab	Graphics
chemistry	Mechanics	Object oriented programming lab	Civil and mechanical
Technical communication and social science	Discrete computational structers	Object oriented programming	Electrical engineering and electronics
System programming	Operational research	Microprocessor	Computer programming
	autometa	Data structure lab	Electrical and mechanical workshop
		Pc hardware and microprocessor lab	Electrical technology
		Mini project	Electronic circuit and logic design
		System programming	Computer organisation
			Logic design lab
			Data structure
			Datacommunication
			Software engineering
			Computer graphics and animation

			dbms
			Knowledge engineering

Final table:-

STUDENT NAME	THEORITICAL SUBJECTS	TECHNICAL SUBJECTS	LOGICAL SUBJECTS	PROGRAMMING SUBJECTS
S1	345	222	214	500
S2				

Fig: Mean marks scored by students

This mean mark in the above table is given as an input to the program. So using mean marks of student we will create a DataSet in which each element represent student and his/her mean marks. And also we create 4 clusters namely theoretical, logical, programming and technical. Then we create a function for k-means algorithm and call this function by passing the dataset as parameter.

First we initialize 4 clusters by randomly generating 4 points(students and their mean mark) and these points are initial cluster centroids. These points are generating by finding the 'minimum instance' and 'maximum instance' of this dataset. For each index the lowest value found in the dataset is the minimum instance and for each index the highest value found in the dataset is the maximum value.Then we find the difference between

each points of corresponding indexes and multiplying it with an random number.

After finding initial centroids we calculate the distance between cluster centroids to each inputs in the dataset and allocate the students to the cluster where the distance from the centroid is minimum. Let us use Euclidean distance, that is,

sqrt((x2-x1)^2 +(y2-y1)^2 + (p2-p1)^2 + (q2-q1)^2)

Knowing the members of each cluster, we compute the new centroid of each cluster based on these new students in cluster. Eg:-

X=(x1+x2+x3+x4)/4

Y=(x1+y2+y3+y4)/4

P=(p1+p2+p3+p4)/4

Q=(q1+q2+q3+q4)/4

Thus new centroids are (X,Y,P,Q)

Again we compute distance between cluster centroids to each inputs in the dataset and allocate the students to the cluster where the distance from the centroid is minimum and compute the new centroid of each cluster based on these new students in cluster .This process continue till that the students does not move clusters anymore. Thus, the computation of the

k-mean clustering has reached its stability and no more iteration is needed.