
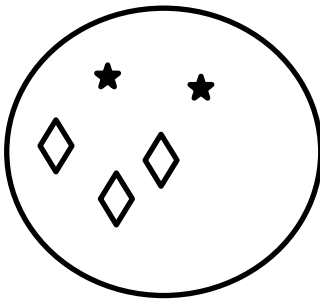


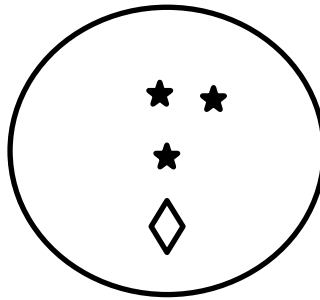
National University of Computer and Emerging Sciences, Lahore Campus

	Course:	Information Retrieval	Course Code:	CS317
	Program:	BS(Computer Science)	Semester:	Fall 2018
	Duration:	25 Minutes	Total Marks:	10
	Paper Date:	30-Nov-18	Weight	3.3%
	Section:	A	Page(s):	2
	Exam:	Quiz 3	Roll No:	

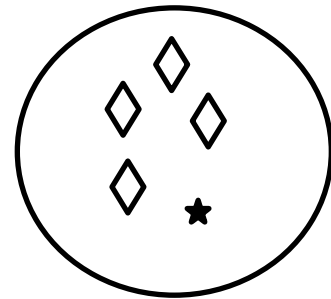
Q1) Compute normalized mutual information of following clusters with classes. There are 2 classes of data. [7 Marks]



Cluster 1



Cluster 2



Cluster 3

Solution:

$$H(\text{cluster}) = \frac{5}{14} * \lg(14/5) + \frac{4}{14} * \lg(14/4) + \frac{5}{14} * \lg(14/5) = 1.57$$

$$H(\text{class}) = \frac{8}{14} * \lg(14/8) + \frac{6}{14} * \lg(14/6) = 0.98$$

$$\begin{aligned} I(\text{Class, cluster}) &= \left(\frac{3}{14} \lg \left(\frac{(14*3)}{(5*8)} \right) + \frac{2}{14} \lg \left(\frac{(14*2)}{(5*6)} \right) + \right. \\ &\quad \left. \frac{1}{14} \lg \left(\frac{(14*1)}{(4*8)} \right) + \frac{3}{14} \lg \left(\frac{(14*3)}{(4*6)} \right) + \frac{4}{14} \lg \left(\frac{(14*4)}{(5*8)} \right) + \right. \\ &\quad \left. \frac{1}{14} \lg \left(\frac{(14*1)}{(5*6)} \right) \right) \\ &= 0.15 \end{aligned}$$

$$NMI = 0.15 / ((1.57) * (0.98)) = 0.097$$

Name _____
Section _____

Roll No _____

Q2) What is RSS value in K Means clustering? Can we use RSS value for determining good value of K in K Means algorithm? Justify your answer. [3 Marks]

Solution:

RSS is residual sum of squares. It is sum of distance of each object with centroid of its cluster.

RSS should not be used for deciding K since RSS decreases as K increases and RSS is 0 for $K = N$ where N is total objects.