
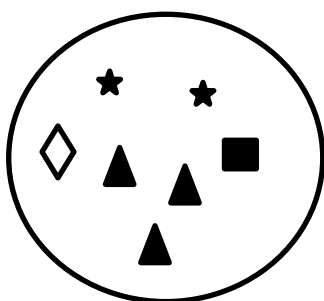


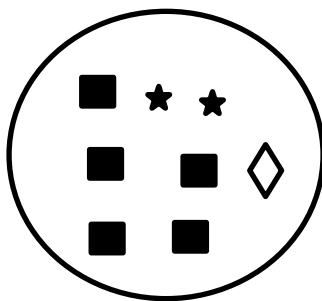
National University of Computer and Emerging Sciences, Lahore Campus

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

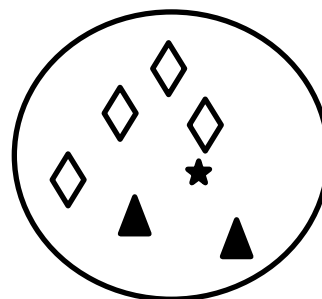
Q1) Compute rand index of following clusters. There are 4 classes of data. [7 Marks]



Cluster 1



Cluster 2



Cluster 3

Solution:

$$\text{All pairs} = 22 \text{ choose } 2 = \frac{22!}{(2! * 20!)} = \frac{(22 * 21)}{2} = 231$$

$$\text{TP} + \text{FP} = (7 \text{ choose } 2) + (8 \text{ choose } 2) + (7 \text{ choose } 2) = 21 + 28 + 21 = 70$$

$$\text{TP} = (3 + 1) + (10 + 1) + (6 + 1) = 22$$

$$\text{FP} = 70 - 22 = 48$$

$$\text{FN} = (3*2) + (1 + 4 + 4) + (4 + 2 + 2) + (5) = 6 + 9 + 8 + 5 = 28$$

$$\text{TN} = 231 - 28 - 70 = 133$$

$$\text{RI} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} = \frac{(22 + 133)}{231} = 0.67$$

Name _____
Section _____

Roll No _____

Q2) If a document has same minimum distance with more than one centroid then it is called a tie. Should we assign the document to any of one the centroid? Why is it important to break ties consistently in K means? [3 Marks]

Solution:

No, we should break ties consistently. Otherwise the document might be assigned to different cluster in each iteration and the algorithm might never converge.