|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **National University of Computer and Emerging Sciences, Lahore Campus** | | | | |
| C:\Users\saif\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Word\final design.jpg | **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:** | **B** | **Page(s):** | **2** |
| **Exam:** | **Quiz 3** | **Roll No:** |  |

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

Cluster 1 Cluster 2 Cluster 3

**Solution:**

All pairs = 22 choose 2 = 22! / (2! \* 20!) = (22 \* 21) / 2 = 231

TP + FP = (7 choose 2) + (8 choose 2) + (7 choose 2) = 21 + 28 + 21 = 70

TP = (3 + 1 ) + (10 + 1) + (6 + 1) = 22

FP = 70 – 22 = 48

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

TN = 231 – 28 – 70 = 133

RI = (TP + TN) / (TP + FP + TN + FN) = (22 + 133) / 231 = 0.67

**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.