|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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:** | **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 (cluster) = 5/14 \* lg (14/5) + 4/14 \* lg(14/4) + 5/14 \* lg(14/5) = 1.57

H (class) = 8/14 \* lg (14/8) + 6/14 \* lg(14/6) = 0.98

I (Class, cluster) = (3/14) lg ((14\*3) / (5\*8)) + (2/14) lg ((14\*2) / (5\*6)) + (1/14) lg ((14\*1) / (4\*8))

+ (3/14) lg ((14\*3) / (4\*6)) + (4/14) lg ((14\*4) / (5\*8)) + (1/14) lg ((14\*1) / (5\*6))

= 0.15

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

**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.