Name _____ Roll No _____ Section _____

National University of Computer and Emerging Sciences, Lahore Campus

Course:	Information Retrieval	Course Code:	CS317
Program:	BS(Computer Science)	Semester:	Fall 2019
Duration:	20 Minutes	Total Marks:	7
Paper Date:	19-Nov-19	Weight	4%
Section:	A	Page(s):	2
Exam:	Quiz 4	Roll No:	

Question 1:

Training	Doc	Words	Class
	1	film dislike unbelievable film	Negative
	2	film comedy greatest awesome film	Positive
	3	end action surprise action enjoy	Positive
	4	pathetic satire movie pathetic	Negative
Test	5	greatest action pathetic comedy film pathetic	**?**

Calculate probability of test document to belong to "Positive" and "Negative" class using **Binarized** Multinomial Naïve Bayes (with Laplace smoothing). Which class will the Naïve Bayes classifier predict for this test document? [5 Marks]

Solution:

After clipping counts at 1 for each document new training and test set

Training	Doc	Words	Class
	1	film dislike unbelievable	Negative
	2	film comedy greatest awesome	Positive
	3	end action surprise enjoy	Positive
	4	pathetic satire movie	Negative
Test	5	greatest action pathetic comedy film	**?**

|V| = 13

Prob("Positive") = 2/4 = 1 / 2

Prob("Negative") = 2/4 = 1/2

Prob (greatest | "Positive") = (1+1)/(8+13) = 2 / 21

Prob (action | "Positive") = (1+1)/(8+ 13) = 2 / 21

Prob (pathetic | "Positive") = (0+1)/(8+ 13) = 1 / 21

Prob (comedy | “Positive”) = (1+1)/(8+ 13) = 2 / 21

Prob (film | “Positive”) = (1+1)/(8+ 13) = 2 / 21

Prob(Doc5 | “Positive”) = (1/2) * (2/21) *(2/21) *(1/21) *(2/21) *(2/21) = 1.95e-6

Prob (greatest | “Negative”) = (1+1)/(6+13) = 1 / 19

Prob (action | “Negative”) = (1+1)/(6+ 13) = 1 / 19

Prob (pathetic | “Negative”) = (0+1)/(6+ 13) = 2 / 19

Prob (comedy | “Negative”) = (1+1)/(6+ 13) = 1 / 19

Prob (film | “Negative”) = (1+1)/(6+ 13) = 2 / 19

Prob(Doc5 | “Negative”) = (1/2) * (1/19) *(1/19) *(2/19) *(1/19) *(2/19) = 8.07e-7

Classifier will predict “Positive” class

Question 2: [2 Marks]

Write any two applications of clustering? One of those applications should be in information retrieval.

Solution

1) Visualization of large collection of documents based on topics of clusters
2) Recall of search engine can be increased by retrieving documents belonging to same cluster of top ranked documents
3) Search engine results can be presented after clustering to resolve ambiguous queries
4) Google news (Presenting crawled news from various sites by clustering based on same news)