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:	B	Page(s):	2
Exam:	Quiz 4	Roll No:	

Question 1:

Training	Doc	Words	Class
	1	Beautiful painting price	NotSpam
	2	Fake painting sale fake	Spam
	3	Great art many replica art great	NotSpam
	4	Replica art value price fake	Spam
Test	5	Replica great art fake money	**?**

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

Solution:

|V| = 11

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

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

Prob (Replica | "Spam") = (1+1)/(9+11) = 2 / 20

Prob (great | "Spam") = (0+1)/(9+11) = 1 / 20

Prob (art | "Spam") = (1+1)/(9+11) = 2 / 20

Prob (fake | "Spam") = (2+1)/(9+11) = 3 / 20

Prob (money | "Spam") = (0+1)/(9+11) = 1 / 20

Prob(Doc5 | "Spam") = (1/2) * (2/20) *(1/20) *(2/20) *(3/20) *(1/20) = 0.000001875

Prob (Replica | "NotSpam") = (1+1)/(9+11) = 2 / 20

Prob (great | "NotSpam") = (2+1)/(9+11) = 3 / 20

Prob (art | “NotSpam”) = (2+1)/(9+11) = 3 / 20

Prob (fake | “NotSpam”) = (0+1)/(9+11) = 1 / 20

Prob (money | “NotSpam”) = (0+1)/(9+11) = 1 / 20

Prob(Doc5 | “NotSpam”) = (1/2) * (2/20) *(3/20) *(3/20) *(1/20) *(1/20) = 0.0000028125

Predicted Class = Not Spam

Question 2: [2 Marks]

(a) What is time complexity of Naïve HAC algorithm?

$O(n^3)$

(b) What is time complexity of Efficient HAC algorithm?

$O(n^2 \lg n)$