CS 5751 – Spring 2018 – Homework 8
Assigned: 04/03/2018
Due: 04/10/2018
Total points: 100 pts.
Submit a soft copy to canvas. Remember to write your name at the top of each file you submit.

Objectives: The objectives of this homework are the following:

- Learn how to fit Naïve Bayes models.

Notes:

- This homework is to be done individually. You may discuss with your classmates, but the work that you write must be your own.

Activity 1: (50 pts.) (Naïve Bayes) Consider the following dataset:

id	X_1	X_2	X_3	Class
1	0	0	1	**1**
2	0	0	1	**0**
3	0	1	1	**0**
4	0	1	1	**0**
5	1	0	1	**1**
6	1	0	1	**0**
7	1	0	1	**0**
8	0	0	0	**1**
9	1	1	1	**1**
10	1	0	1	**1**

a) (10 pts.) Compute P(A = 1 | Class = 1), P(B = 1 | Class = 1), P(C = 1 | Class = 1), P(A = 1 | Class = 0), P(B = 1 | Class = 0), P(C = 1 | Class = 0). Show your work step by step.
b) (10 pts.) Classify the example (A = 0, B = 1, C = 0) using the Naïve Bayes approach. Show your work step by step.
c) (10 pts.) Estimate the conditional probabilities using the m-estimate approach (check the slides).
d) (10 pts.) Repeat part (b) using the conditional probabilities you found in part (c).
e) (10 pts.) Compare the two methods for estimating probabilities. Which method is better and why?

Activity 2: (50 pts.) (Naïve Bayes) Using Python or R, train a Naïve Bayes model using the dataset of Activity 1, then classify the same example (A = 0, B = 1, C =0). Verify if your answer is the same as in Activity 1b or as in Activity 1c.

For this activity, write a Jupyter notebook named yourLastName_hw8.ipynb.