CS 5751 – Spring 2018 – Homework 9
Assigned: 04/10/2018
Due: 04/17/2018
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 about ROC curves.

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: (100 pts.) (ROC Curves) Consider the same dataset as in last week's homework:

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**

And also consider a classification model M' (it could be something like a logistic regression classifier, but we don't really care) that outputs the following probabilities for each of the data points of the above dataset:

id	$P(Class = 1 \mid X_1, X_2, X_3, M')$
1	**0.73**
2	**0.69**
3	**0.44**
4	**0.55**
5	**0.67**
6	**0.47**
7	**0.08**
8	**0.15**
9	**0.45**
10	**0.35**

a) (10 pts.) Fit a Naïve Bayes classifier to predict Class 1 (the class we are interested in) using the m-estimate approach. For this task, you can reuse your computations from last week.
b) (30 pts.) In the same figure, plot the ROC curve for the Naïve Bayes classifier of part 1a, plot the ROC curve for the M' classification model, and plot the ROC curve for the random guesser. Your plot must have a caption, a legend, and its axes must have proper labels. You must show the calculations that you used to plot the curves. You can check your answers with sklearn or caret.
c) (20 pts.) Compute the AUC for all three classification models.
d) (20 pts.) Which of all three models do you think is better? Explain your reasons.
e) (20 pts.) For the Naïve Bayes model, choose a cutoff threshold of 0.6. Then compute the precision, recall and F-measure for the model at this threshold value.

For this activity, submit a PDF file containing your answers. If you also submit code, then submit it as a Jupyter notebook.