Machine Learning
Georgia State University
Professor: Jaya Krishna Mandivarapu

Home Work 4

Submission Requirements

You must turn work at the SPECIFIED TIME so you can receive credit for Homework!

Files Required for submission : One Jupyter Notebook and HTML file (Can be downloaded from Jupyter notebook you are working with) and word document incase if you prefer to write theoretical answers in word document

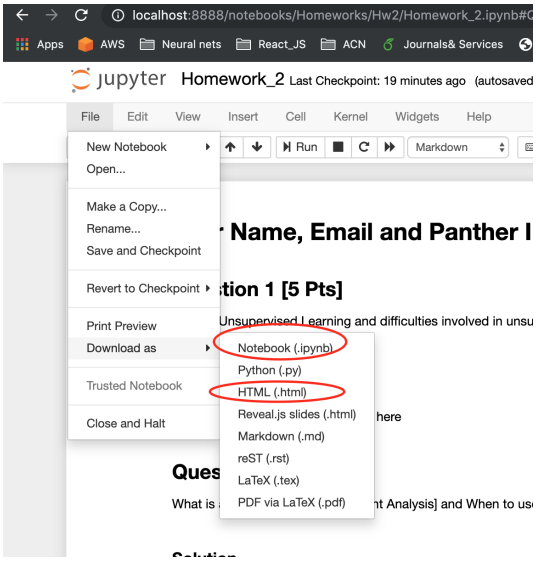

Figure 1: Download as Jupyter Notebook and HTML file

Homework 4must be **submitted on icollege** by the due date and time. Late homework will be subject to a penalty of 30 percent for 1 day and 60 percent for two days and after 3 days no submission allowed, as stated in the course grading policy. No email or hard copies of homework will be accepted.

You may discuss the assignments with other students in the class, but (as stated in the academic honesty policy) your written answers **must be your own**, and you must list the names of other students you discussed the assignment with.

How to Submit

Log into **iCollege(iCollege)**, select the class to view its drop box folders, select the correct folder for the given assignment and upload the file there.

You will get a confirmation email. Please save the conformation email in the event something goes wrong, for example work was submitted to the wrong folder etc..

Machine Learning
Georgia State University
Professor: Jaya Krishna Mandivarapu

Home Work 4

1. a) What is Conditional probability, Marginal probability and Joint probability? Write their mathematical formulas and give one example each. [5 pts]

   b) Explain what is Baye's rule with the formula and what is prior, posterior, likelihood and marginal probability in the Baye's rule. [10 pts]

   c) What is Naive Bayes algorithm and how is related or derived or inspired from Bayes rule? [5 pts]

2. Perform Naive Bayes algorithm on the below dataset in python in which you can classify test example weather a **Red Domestic SUV** is stolen or not as shown in 2.2. [10 pts]

   ## 2 Car theft Example

   Attributes are Color , Type , Origin, and the subject, stolen can be either yes or no.

   ### 2.1 data set

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

   ### 2.2 Training example

   We want to classify a Red Domestic SUV. Note there is no example of a Red Domestic SUV in our data

   Hints and Explanation:

   convert 1(Table 1),2.1 into a csv then load a csv file or you can prepare your dataframe from 2.1 dataset. [As i posted the text for 2.1,Table 1]

3. Perform Naive Bayes algorithm on the below problem3.csv dataset in python. [70Pts].
   1. Print out Accuracy Score, Confusion matrix for both training and test set.
   2. Print True Positives(TP),True Negatives(TN), False Positives(FP) ,False Negatives(FN) for both train and test sets.
   3. Visualize confusion matrix with heatmap plot for both train and test sets.
   4. Print the classification report (precision,recall or Sensitivity and f1-score) for both test and train dataset
   5. As the naive bayes algorithm is based on probabilities, please calculate class probabilities of each item in the test set. Then visualize the histogram of class probabilities of each class of target variable separately (Our Target Variable is Income which has two unique classe, so two histograms)

   Hints and Explanation: 1.Perform Exploratory data analysis on the entire dataset (finding missing values or any special characters, converting categorical values to numerical values ..etc,), 2. Don't remove the missing value row instead use the methods

discussed in class like column mean,mode ..etc to impute. 3. Split train and test in the ratio of 70 and 30. 4. For calculating the class probabilites one can use the sklearn module directly.