

Due Date: November 3rd 2019 at 11.59 PM

Submission Requirements

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

Please provide all the work in single html, jupyternotebook file. Please don't submit multiple different files like word,pdf. We only accept two file .html,jupyternotebook file

Files Required for submission : One Jupyter Notebook and HTML file (Can be download from Jupyter notebook you are working with)

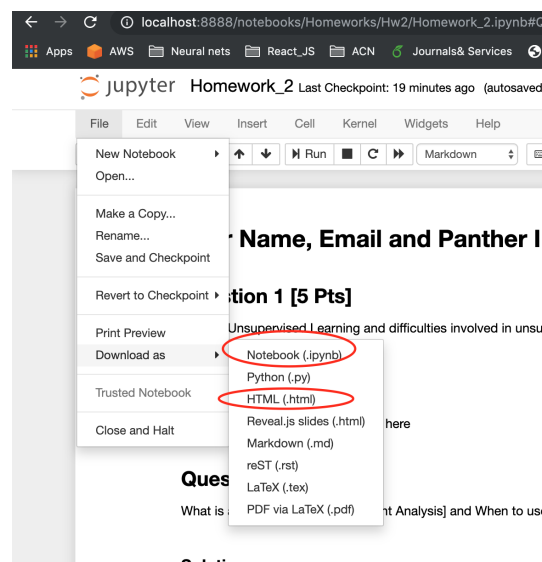


Figure 1: Download as Jupyter Notebook and HTML file

Homework 2 must be **submitted on icollege** by the due date and time. Late homework will be subject to a penalty of 50 percent for 1 day and 80 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..

1. What is cross-validation? Explain everything in detail about how the cross validation can be performed on given dataset[10 pts]
2. What are advantages and disadvantages of Cross validation? [5 Pts]
3. What is Generalized Additive Models?[Please refer to our course textbook Elements of Statistical Learning Textbook for this] [5 Pts]
4. How does Decision Tree Based Method's differ from the rest of the classification methods? [5 Pts]
5. What are different classification techniques available and How does a Decision Tree work? [Please write in atleast 5 sentences] [10 pts].
6. What are some of the advantages and disadvantages of decision trees? [5 pts]
7. What are different subset selection based methods and explain the difference between subset selection based method's vs cross validation [Explain with example] [10 Pts]
8. Please Perform Logistic Classification on the Give dataset (insurance-data.csv) below.[10 Points]

Dataset Link : [Dataset](#)

10 Points for Logistic Regression accuracy and output plots.

9. Please Perform Decision Tree Classification on the Give dataset(data.csv) below.[40 Points]

Dataset Link : [Dataset](#)

15 Points for Data Preprocessing.

15 Points for Decision Tree along with Tree Plot at the end and Results Explanation.

10 Points for applying the crossvalidation on the given data and show all the accuracies for each split. Please do mention what is the best model to use.

Hints:

As per the data preprocessing step convert all the variables in the dataset into Numerical values as the algorithms only work with Numerical values