

**Due Date: Oct 2nd 2019 at 11.59 PM**

## Submission Requirements

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

*Did posted a jupyter notebook which contains place where you need to enter your answers. So, Please download the jupyter notebook to your system then use that to submit your answers*

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

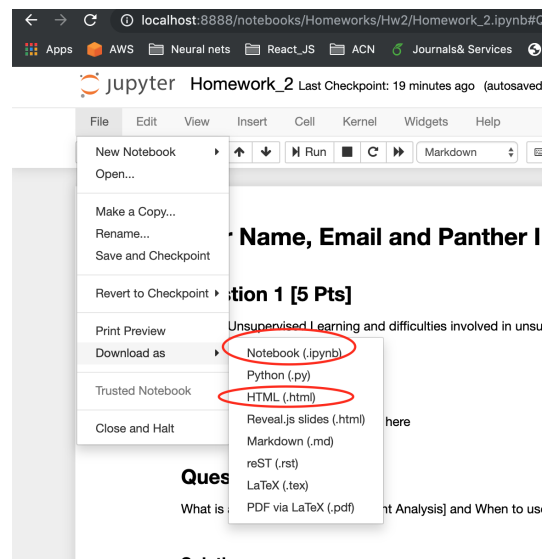


Figure 1: Download as Jupyter Notebook and HTML file

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 Unsupervised Learning and difficulties involved in unsupervised Learning and name a few unsupervised algorithms ? [5 pts]
2. What is a PCA? When to use PCA ? [Please explain with one to two examples] [5 Pts]
3. How does a PCA work? [Please write in atleast 5 sentences] [5 pts].
4. What are different pre-processing steps you can apply to the features before you can compute the PCA? Does every pre-processing method will yield same result or answer at the end of each method? [5 pts]
5. What are advantages and disadvantages of PCA? [Explain with example] [5 Pts]
6. What is clustering? Explain how K-Means Clustering Algorithm works?
7. What are the Advantages and disadvantages of Clustering Algorithms discussed in our class (K-Means, Hierarchical)?
8. Which Clustering Algorithm is better K-Means or hierarchical Clustering? Explain with a proper example which is better algorithm in each scenario?
9. Please Perform Principal Component Analysis, Hierarchical and K-Means Clustering on the Give dataset Below. [50 Points]

Dataset Link : [Dataset](#)

10 Points for Data Preprocessing.

10 Points for PCA Algorithm along with plots and Results Explanation.

10 Points for K-Means Algorithm with plots and Results Explanation.

10 Points for Hierarchical Algorithm with plots and Results Explanation.

10 Points for Comparing the results between PCA, Hierarchical Algorithm and K-Means and what's your inference from your outputs of the algorithms. Can you mention which algorithm works best for clustering on given dataset according to your plots?

**Hints:**

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

Then Apply three algorithms one after the other then plot the output clusters

Compare the output clusters in all the steps.