Machine Learning
Georgia State University
Professor: Jaya Krishna Mandivarapu **Home Work 1,2**

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)

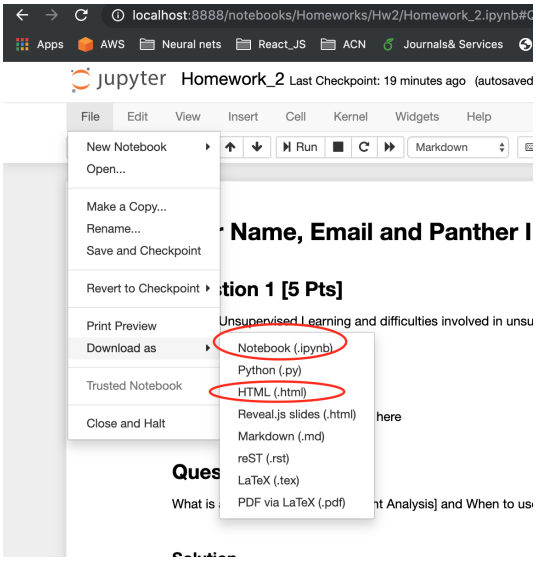

Figure 1: Download as Jupyter Notebook and HTML file

Homework 1,2must 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..

Machine Learning
Georgia State University
Professor: Jaya Krishna Mandivarapu **Home Work 1,2**

1. What is Supervised Learning, unsupervised learning and difficulties involved in unsupervised learning and name a few supervised, unsupervised algorithms ? [5 Pts]

2. What are parametric and non-parametric methods?. Explain pro's and con's of each.[5 pts]

3. Explain about standard error , Co-variance and co-relation. When each one of them is used? [5 Pts]

4. Explain how hypothesis testing is used in determining the linear relationship between the feature and target variable in linear regression? [5 Pts]

5. Please write down the equation of multi linear regression and explain each term present in it? [5 Pts]

6. What is a PCA? When to use PCA ? How does a PCA work? [Please write in atleast 5 sentences] [5 pts].

7. 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]

8. What is clustering? Explain how K-Means Clustering Algorithm works?[5 pts]

9. What are the Advantages and disadvantages of Clustering ALgorithms discussed in our class (K-Means,Hierchal)?[5 pts]

10. Which Clustering Algorithm is better K-Means or hierarchical Clustering? Explain with a proper example which is better algorithm in each scenario? [5 pts]

11. Please Perform Principal Component Analysis, Hierarchical and K-Means Clustering on the Give dataset Below. [50 Points]

    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 whats 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. Check weather you

need to standardization or normalization of data. Then Apply three algorithms one after the other then plot the output clusters

Compare the output clusters in all the steps.

12. Answer the following? [50 pts]

a) Write and explain about the Linear Regression and it's equation [3 pts]

b) Explain in detail about the loss function of linear regression, R^2, Adjusted R^2 used in the Linear Regression and what is the need for Adjusted R^2? [12 pts]

c) Plot X vs Y in Scatter plot from data in Table 1 and comment on the relation of X vs Y using Covariance,Corelation. Please comment on Covariance and corelation values [5 Pts]

d) Perform Linear regression on the following data using Python? and print β_0, β_1 values in equation y= β_0+ β_1*x. Please write down what is your understanding from those values. [10 Pts]

e) What are different evaluation metrics available for predicting the performance of the Linear Regression? Evaluate all those methods on the given dataset in Table 1 and also please print out the accuracy, R^2, Adjusted R^2 [10 pts]

X	Y
6	526
3	421
6	581
9	630
3	412
9	560
6	434
3	443
9	590
6	570
3	346
9	672

Table 1: X(No of Weeks) vs Y(Avg Sales)