

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

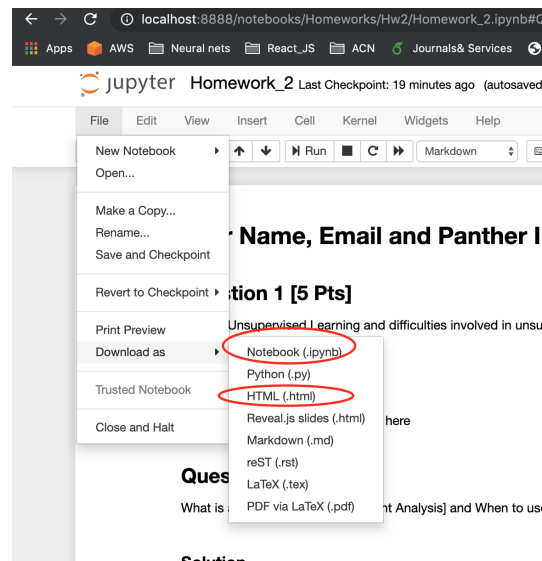


Figure 1: Download as Jupyter Notebook and HTML file

Homework 5 must 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..

1. Perform SVM algorithm on the below **svm.csv** dataset in python.(Final column **target_class** is the target variable) [50Pts].

a) Visualize the pair plots between all the variables and heat map plot of correlation between all the variables

b) Split the original dataset in 70/30 ratio for train and test set then perform Linear Support Vector Classifier(from sklearn.svm import LinearSVC) on the dataset. Print out Accuracy Score, Confusion matrix, Visualize confusion matrix with heatmap plot, classification report (precision,recall or Sensitivity and f1-score) for both training and test set.

C) Split the original dataset in 70/30 ratio for train and test set then perform Support Vector Classifier(from sklearn.svm import SVC) using **polynomial kernel with degree=2** on the dataset. Print out Accuracy Score, Confusion matrix, Visualize confusion matrix with heatmap plot, classification report (precision,recall or Sensitivity and f1-score) for both training and test set.

d) Split the original dataset in 70/30 ratio for train and test set then perform Support Vector Classifier(from sklearn.svm import SVC) using **rbf kernel with gamma=0.5, C=0.1** on the dataset. Print out Accuracy Score, Confusion matrix, Visualize confusion matrix with heatmap plot, classification report (precision,recall or Sensitivity and f1-score) for both training and test set.

e) Please comment on outputs of problems b,c,d.(Like which one performing best and why and so on.)

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.

2. Perform Decision Tree algorithm on the below **decisionTree.csv** dataset in python.(Final column **target_class** is the target variable) [50Pts].

a) Split the original dataset in 70/30 ratio for train and test set then perform Decision Tree algorithm (from sklearn.tree import DecisionTreeClassifier) with **criterion as gini, max_depth=3, random_state=0**. Print the train and test set accuracies and visualize the decision tree.

b) Split the original dataset in 70/30 ratio for train and test set then perform Decision Tree algorithm (from sklearn.tree import DecisionTreeClassifier) with **criterion as**

Machine Learning

Georgia State University

Professor: Jaya Krishna Mandivarapu **Home Work 5**

`entropy, max_depth=3, random_state=0`. Print the train and test set accuracies and visualize the decision tree.