Name: Xinyuan Liu

Student ID:5183584

Email:liux3372@umn.edu

Instructions:

The version of python is 2.7.11

In question 3, Change to the source code directory and run the codes with python q3i.py or python q3ii.py which will import my\_cross\_val.py and my\_train\_test.py, respectively

In question 4, Change to the source code directory and run the codes with python q4.py which will import rand\_proj.py, quad\_proj.py and my\_cross\_val.py

We may need to reinstall the sklearn package if there is an import error with sklearn.model.selection:

pip install sklearn

The output for each run will give the method dealing with certain dataset (e.g. LinearSVC with Boston50) and a table with 12 columns (with labels of col0 through col11) which contain every cross-validation run in a total of 10 times (i.e. K=0, K=1, … K=9) followed with the mean and variance of these 10 runs.

The tables in the pdf are reorganized to be more readable. I only create a documented table and change the header to be K=0, K=1, … K=9 following with Mean and Std (i.e. Standard Deviation).

The arguments for different methods in both my\_cross\_val and my\_train\_test are the string of the method names (i.e. "LinearSVC", "SVC" and "LogisticRegression")

It may encounter a DeprecationWarning: using a non-integer number instead of an integer will result in an error in the future, that is because indexing with non-integer in scikit-learn. It won't affect the results.