Lab 5

Implement Linear Regression and Ridge Regression in Python. Please do not use any machine learning library.

Training data: http://www.cse.scu.edu/~yfang/coen140/crime-train.txt
Test data: http://www.cse.scu.edu/~yfang/coen140/crime-test.txt
A description of the variables: http://www.cse.scu.edu/~yfang/coen140/communities.names

The data consist of local crime statistics for 1,994 US communities. The response y is the crime rate. The name of the response variable is *ViolentCrimesPerPop*, and it is held in the first column of *df_train* and *df_test*. There are 95 features x_i. These features include possibly relevant variables such as the size of the police force or the percentage of children that graduate high school. The data have been split for you into a training and test set with 1,595 and 399 entries, respectively. The features have been standardized to have mean 0 and variance 1.

Exercises:

- Implement the linear regression model. Compute the RMSE value on the training data and test data, respectively. Report both RSME values.
    - Provide a function that, given an NxP numpy array of data points, returns an Nx1 array of predicted outputs.
    - Define this function as follows:
        def problem1(samples):

- Implement the ridge regression model with $\lambda = 100$. Compute the RMSE value on the training data and test data, respectively. Report both RSME values.
    - Provide a function that, given an NxP numpy array of data, returns an Nx1 array of predicted outputs.
    - Define this function as follows:
        def problem2(samples):

Notes:

*** Note that N = number of instances, P = number of features.

*** Submit in the form of a Jupyter notebook. Recall that you can report your 4 RSME in a Markdown cell within the notebook—you do not need to print these out within the above functions.

*** Use the datasets provided without modifications. Do not rename or alter the file's contents.