Due: 4/9/2022 23:59PM Submit to myCourses

Note:

1. *Providing only the result (e.g., a figure) will get 0 pts*
2. *Please submit your source codes and name them via question numbers*
3. *Please submit your experimental results in a single document*
4. *You are allowed to use MATLAB or Python, or other languages for experiments*

Q1. (40 pts) In this question, you will explore two different regression models and make a head-to-head comparison. Please use **Abalone** dataset in the folder for this task. There are 8 different measurements as input features, and 1 output (ages), and the regression model will predict the age of Abalone based on these 8 measurements. Assume the features matrix is A, the **regression parameter vector is $\boldsymbol{\beta}$**, and the age vector is y. A simple linear regression model aims to approach the problem: $A\beta = Y$ where β is unknown. Could you use:

a. (20 pts) Linear regression $\min\|A\beta - Y\|_F^2)$

b. (20 pts) Ridge regression ($\min\|A\beta - Y\|_F^2 + \lambda\|\beta\|_2^2$, namely, adding l_2 norm on β)

to solve this problem (**Hint**: a sample code called "RegressionSample.m" has been included in the folder). In each regression experiment, please show the followings:

Use a ***50/20/30 training/validation/testing split***, and then report the ***average square-error***: $(y' - y)^2/n$ on **test dataset**, where y' is the predicted value, y is the ground truth value, and n is the total number of testing data. In Ridge Regression, you will pick a proper value for the regularization parameter λ (e.g., $10^{-4}, 10^{-3}, 10^{-1}, 1, 10, etc.$) ***through validation set***. Please visualize different β using ***bar graph*** when given different λ.

Q2. (60 pts) PCA and LDA are two important linear dimensionality reduction methods. In this question, you will compare different implementations of PCA and LDA, given CMU PIE dataset. The CMU PIE dataset contains 67 subjects, and 21 images under different lightings for each subject. The size of each image is 30×30.

a. (30 pts) Please implement PCA and LDA by yourself, and then use the provided CMU PIE dataset (PIE.mat) for evaluations. The gist of your implementations can be found in "myPCA" and "myLDA". In this evaluation, we mainly focus on face recognition, and use the "Nearest Neighbor Classifier". Please run with different training sets, i.e., trainNum = 5,10,15 per subject (The rest will be used for testing purpose), and record the running time for each setting. In total, there will be 6 recognition accuracies and 6 running time for PCA and LDA. You are recommended to use a table to ***summarize all the results***.

b. (20 pts) Use the online code 'PCA.m' and 'LDA.m' provided in the package for the same tasks as the last question. Use a table to ***summarize all the results***. (**Hint**: a sample code has been provided in the package, SampleTest.m)

c. (10 pts) Please use the ***first five leading eigenvectors*** (corresponding to the first five largest eigenvalues) from PCA to visualize EigenFace. Basically, you will have five EigenFaces shown in your solutions. (**Hint**: you need to convert the eigenvector back to a 30×30 matrix, same size as face images, followed by some pixel value (scale) adjustment to make sure they are shown properly.)

[provided files list]: PCA.m, LDA.m, myPCA.m, myLDA.m, LGE.m, mySVD.m, SampleTest.m, PIE.mat