CSE 404: Introduction to Machine Learning (Fall 2019)

Homework #10

Due 12/4/2019 before class

1. (75 points) This question is related to Principle Component Analysis (PCA). You are to apply data pre-processing techniques to a collection of handwritten digit images from the USPS dataset. You can load the whole dataset into Python using the function `loadmat` in `Scipy.io`. The matrix A contains all the images of size 16 by 16. Each of the 3000 rows in A corresponds to the image of one handwritten digit (between 0 and 9).

   (a) (25 points) Implement PCA

   (b) (25 points) Apply PCA to the data using $d = 10, 50, 100, 200$ principal components.

   (c) (25 points) Reconstruct images using the selected principal components from part 1.

   Your submission should consists of the following

   - The source codes for parts 1, 2 and 3.
   - The total reconstruction error for $d = 10, 50, 100, 200$
   - A subset (the first two) of the reconstructed images for $d = 10, 50, 100, 200$.

   NOTE: You are NOT supposed to use existing PCA code; instead, you should write your own PCA function using eigen-decomposition (you can use existing packages for eigen-decomposition.).