ECE 493/579: Machine Learning for IoT
Department of Electrical and Computer Engineering
Rutgers University

RUTGERS
School of Engineering
ELECTRICAL AND COMPUTER ENGINEERING

ECE 493/579 Machine Learning for the Internet of Things (Fall 2022)
Homework #2 (120 points)

Rationale and learning expectations: This homework covers some of the basic topics from linear algebra, which are necessary for understanding of most concepts in machine learning. All students, regardless of their targeted grade in this course, are expected to have complete mastery of the concepts being tested in this homework.

1. **PCA from Scratch. [30 points]** Given the following pseudocode for Principal Components Analysis (PCA). Implement it in code and run it on the following faces dataset.

   a) *Normalize/Standardize the data*. The first step will be to normalize or standardize the data. That is, you want to maintain a similarly small range for all features in your samples. Note, I recommend normalizing it since PCA optimizes for covariance across features. **[5 points]**

   b) *Center the data*. Calculate the mean of the data and remove it from each samples. (The mean of the centered data should be equal to 0).**[5 points]**

   c) *Calculate the covariance matrix*. Compute the covariance matrix for your data. **[5 points]**

   d) *Calculate Eigenvectors and Eigenvalues of the covariance matrix*. Use the linear algebra package to perform a decomposition of your covariance matrix into a product of eigenvectors and eigenvalues. **[5 points]**

   e) *Choose top k principal components* (top k eigenvectors w/associated eigenvalues). Sort the eigenvectors in descending order corresponding to their associated eigenvalues. **[5 points]**

   f) *Project the data* to principal component axes and plot it across the top two principal components. **[5 points]**

2. **Unroll PCA [30 points]** Write a function that takes each of the points in your data in PCA-space and projects them back to dataspace. Write a function that compares the two data sets and compute the average difference between them. Print out the result. *Hint: this requires keeping track of the transformations applied to the data an applying their inverse in reverse order.*

3. Use the faces dataset to compare the distance matrix in pixel space versus the distance matrix in PCA space (projected space). **[30 points]**

   a) What would the ideal fit be? **[2 points]**

   b) Compute the average error? **[8 points]**

   c) What is the error in comparison to ideal? **[12 points]**

   d) Show a plot that compares them. **[8 points]**

4. Use the faces dataset to compare the distance in pixel space versus those in projection space using MDS. You may use the MDS library from scikit-learn. **[30 points]**

a) What would the ideal fit be? **[2 points]**

b) Compute the average error? **[8 points]**

c) What is the error in comparison to ideal? **[12 points]**

d) Show a plot that compares them. **[8 points]**