

Assignment-2*Instructor:* Prof. M. N. Murty*TAs:* Kishalay Das, Paarth Gupta**Assignment Policy:**

Welcome to the Second Assignment of Linear Algebra, where you will implement Spectral Clustering and SVD. **Read all the instructions below carefully before you start working on the assignment, and before you make a submission.**

- This is a coding assignment and the code has to be written in Python. Use Python version 3 to solve all the problems. The python file name should be **main.py**.
- **For computational part you are not supposed to use any library function. Do code from scratch.**
- The assignment deliverable is your python code (main.py) and your report(pdf format).
- Send above files in a zip file to lap_e0226@outlook.com with the following format for both zip file name and the subject name of email: **LAP_2_XXXXX**(Last 5 digits of your Student ID)
- **Read the complete assignment carefully, before attempting to solve it.**
- **For each problem report the results of experiment you have done for the given train dataset. Also submit the code which we will verify for test dataset.**
- You are encouraged to discuss among yourselves, but **DO NOT COPY** solutions or code. **The consequences will be severe.**
- The submission deadline is **24th November, 2019 11:59 P.M..**

El-Clásico:The Greatest Soccer Derby!

El Clásico (or "The Classic") is the greatest football rivalry between fierce rivals Real Madrid and FC Barcelona. Originally it referred only to those competitions held in the Spanish championship (La-Liga), but nowadays the term has been generalized, and tends to include every single match between the two clubs: UEFA Champions League, Copa del Rey, etc. Moreover for the last one decade this rivalry got extra significance due to two of the greatest players of all time playing against each other in this match: Lionel Messi and Cristiano Ronaldo. Hence, it is considered as one of the biggest club football games in the world, and is among the most viewed annual sporting events.

Outside the football field rivalry, players of both the football clubs share good relationship with each other. Players from both the clubs frequently go in a sports pub at Barcelona, named Flaherty's Irish Pub. And once they have been in the pub they interact with other players (of their own or rival club) and enjoy their evening. The manager of the pub has collected data of those pair of players who interacted inside the pub, since the start of the La-liga season 2019-20. From all such links he formed a network and now he wants to find the two clusters present in that network. Basically the two clusters will represent how closely group of players interacts with each others.

Suppose the manager hires you to solve this problem. You being a Linear Algebra enthusiast, has to propose the manager of the pub an algorithm to help him finding two clusters.

Problem 1

(1.25 points)

Find the Eigen value and Eigen vectors of any given matrix. **You have to write the code for eigen vector and values from scratch without using any library function.** Report the algorithm you have used.

Query: python main.py problem1 input.txt

Expected Output: Eigen Values and Vectors of the Matrix

Sample Input: ([Sample file here under problem1 folder](#))

Problem 2

(1.25 points)

Find centrality measure of each node of the football network (mentioned above) using **Degree Centrality, Closeness Centrality, Betweenness Centrality, Eigenvector Centrality** measure. Report all the results you observed in different measures. Also submit the code for each centrality measure. **Use networkx package to read the gml file. But for the centrality measures implement your own algorithm without using any library function.**

Query: python main.py problem2 test-network.gml

Expected Output: Centrality measure of all the nodes of the test network by each of the centrality measure

Input Network: ([Sample gml file](#))

Problem 3

(1.25 points)

Find two clusters in the football network using spectral clustering algorithm. While implementing the algorithm use code of problem1 to find eigen value and vector for the Adjacency matrix of the network. **Plot the network (using any library function) with separate colors for the nodes of each cluster.** Report the plot for the train network dataset. **Implement the algorithm from scratch without using any library function.**

Query: python main.py problem3 test-network.gml

Expected Output: Plot for two clusters of the test network dataset.

Input: Your code will be evaluated with a test network dataset.

MNIST Handwritten Dataset

Dataset description: MNIST is a handwritten dataset, originally has 60,000 digits with 784 (28 X 28)

dimension in its training set. It has 10,000 X 784 in its test set. In this assignment we use a subset of the dataset of size 10,000 X 784 for training and 2,000 X 784 for testing if required. (Note: In the provided files first column has digit labels. Ignore the label column when you reconstruct the matrix)

Problem 4a

(1 points)

Given a subset of MNIST (mnist-train.csv) dataset your goal here is to find the truncated SVD for different values of d . Take a dataset $X \in R_{m,n}$ and compute SVD for $d = 2, d = 5, d = 10, d = 20, d = 50, d = 100, d = 200, d = 500$ and plot reconstruction error $RMSE(X, \hat{X}) = \sqrt{\sum_{i,j} (X_{i,j} - \hat{X}_{i,j})^2}$ for the train dataset.

Use Matplotlib to plot RMSEs. But SVD algorithm you have to implement by your own from scratch without using any library function. But for finding eigen values and vectors in this problem use library function.

Query: python main.py problem4a mnist-test.csv

Expected Output: Plot the RMSE.

Input: MNIST test dataset. ([MNIST train dataset](#))

Problem 4b

(0.25 points)

Project the vectors into 2D space and plot them (Using tSNE). Do you see any pattern in the plot? Report the plot for the train data.

Query: python main.py problem4b mnist-test.csv

Expected Output: Plot the 2D space projection using tSNE for test dataset.

Input: MNIST test dataset.