

CSECE 8735 Fall 2023  
Unsupervised Learning

**Assignment 3**  
due Thursday 10/26/2023

**Note: please use two digits (apply rounding) after the decimal point when entering your answers.**

**Problem 1** Carry out latent semantic analysis on the word-document matrix  $W$  (12 words, 10 documents) given in WordDocF23.dat. (c.f. class notes of 10/17/23)

- 1.1 Perform SVD decomposition on  $W$  (use `svd()` in MATLAB, choose economy size).
  - 1.1.1 Determine the largest singular value
  - 1.1.2 Based on the relation between SVD and PCA, determine the largest eigenvalue from PCA for  $W$ .
- 1.2 Perform rank- $k$  approximation on  $W$  for  $k = 1, 3, 5$  with the approximation error defined by  $\|X - \hat{X}\|_F^2$ .
  - 1.2.1 What is the approximation error for  $k = 1$ ?
  - 1.2.2 What is the approximation error for  $k = 3$ ?
  - 1.2.3 What is the approximation error for  $k = 5$ ?
- 1.3 Based on the rank-3 approximation on  $W$ , determine the following inner-product similarity terms:
  - 1.3.1 The most similar document pair and the similarity value.
  - 1.3.2 The most similar word pair and the similarity value.
- 1.4 Based on the rank-3 approximation on  $W$ , determine the following cosine similarity terms:
  - 1.4.1 The most similar document pair and the similarity value.
  - 1.4.2 The most similar word pair and the similarity value.

**Problem 2** Use the classical multidimensional scaling algorithm (MDS) to estimate the 2-D coordinates of nine cities in Missouri: Branson, Cape Girardeau, Columbia, Jefferson City, Kansas City, Rolla, St. Louis, Springfield, and St. Joseph (the data points  $x_1$  through  $x_9$  are arranged in this order). The Googled driving distance between each pair of the nine cities (in miles) are given in MOCityDistF23.dat.

(Note: please keep the first components of the two eigenvectors in  $V$  + negative; if you need to change the sign of an eigenvector, multiply the whole vector by -1.)

- 2.1 What is the largest eigenvalue?
- 2.2 What is the second largest eigenvalue?
- 2.3 Determine the 2D coordinate of Columbia ( $x_3 = [x_{31} \ x_{32}]$ ).
- 2.4 Determine the 2D coordinate of Rolla ( $x_6 = [x_{61} \ x_{62}]$ ).

2.5 Based on the estimated coordinates of the data points, re-compute the following Euclidean distance terms:

2.5.1 Recomputed Euclidean distance between Columbia and St. Joseph;

2.5.2 Recomputed Euclidean distance between Kansas City and Springfield.

2.6 Plot a map for the nine cities based on their MDS reconstructed coordinates (use a scatter plot, and print the name of each city alongside the point of the city). Upload your map in pdf format.