CSEN 240
Homework #7

(Level 1: high-level understanding)

1. (12 points) Which of the following statements about PCA (Principal Component Analysis), SVD (Singular Value Decomposition), and eigenvalue decomposition (EVD) is/are true? Select all that apply.
    a. PCA is a supervised learning algorithm.
    b. The principal components can be obtained via via EVD on $\mathbf{X}\mathbf{X}^T$ (where $\mathbf{X}$ is the training dataset). They are linear combinations of the original variables.
    c. The first principal component captures the most variation in the data.
    d. SVD is a factorization of a matrix into three matrices.
    e. Left singular vectors of $\mathbf{X}$ are the same as eigenvectors of $\mathbf{X}\mathbf{X}^T$
    f. EVD can be applied to non-square matrices.
2. (10 points) Which of the following statements are true about Linear Discriminant Analysis (LDA)? (Select all that apply.)
    a. LDA is a supervised dimensionality reduction technique used for feature extraction.
    b. LDA aims to find a linear combination of features that maximizes the separation between different classes.
    c. The Fisher criterion aims to maximize the ratio of between-class variance to within-class variance.
    d. LDA can only be applied to binary classification problems and is not suitable for multiclass problems.
    e. The output of LDA includes the projection of data points onto the extracted discriminant features.

(Level 2: manual exercise)

3. (10 points) PCA
    a. Given the following dataset of unlabeled points in two dimensions
        i. { (1, -2), (2, -1), (3, -3), (4, -4) }
    b. To perform PCA on the dataset, please
        i. Preprocess the dataset, center it around zero by calculating the dataset's mean and adjusting each data point relative to the mean, and form the new dataset into a matrix **X**
        ii. Calculate $\mathbf{X}\mathbf{X}^T$
        iii. Perform eigenvalue decomposition on $\mathbf{X}\mathbf{X}^T$ (Please check the lecture note on perform EVD by hand)
        iv. Calculate the variance of the dataset projected on the first principal component and the variance of the dataset projected on the second principal component
        v. What do you observe?
4. (18 points) LDA
    a. Given the following dataset of labeled points in two dimensions

i. Class 1: { (-1, 2), (-2, 1) }
ii. Class 2: { (-3, 3), (-4, 4) }

b. To perform LDA on the dataset, please
   i. Preprocess the dataset, center it around zero by calculating the dataset's mean and adjusting each data point relative to the mean, and form the new dataset into a matrix **X**
   ii. Calculate mean for each class
   iii. Compute the within-class variance matrix $\mathbf{S}_W$
   iv. Compute the between-class variance matrix $\mathbf{S}_B$
   v. Perform eigenvalue decomposition on $\mathbf{S}_W^{-1}\,\mathbf{S}_B$
   vi. Calculate the within-class and between-class variance of the dataset projected on the first component
   vii. Calculate the within-class and between-class variance of the dataset projected on the second component

(Level 3: computer-based exercise)

5. (20 points) PCA (Using HW7-5.ipynb)
   a. Load the iris dataset and remove the first class
   b. Implement PCA by yourself
   c. Measure the accuracies using from 1 to 4 principal components (10 times each)
   d. Report the average accuracy and your observation
6. (20 points) LDA (Using HW7-6.ipynb)
   a. Load the iris dataset and remove the first class
   b. Implement 2-class LDA by yourself
   c. Measure the accuracies using from 1 to 4 principal components (10 times each)
   d. Report the average accuracy and your observation