AI 221: Machine Exercise 4

Instructions:

- Read and answer each problem using computer code. This MEX should be done *individually*.
- Each item should be answered as either a Jupyter Notebook or a MATLAB Live Script, exported as a single PDF file for the entire MEX. Make sure to HIGHLIGHT your final answers.
- When done, submit the PDF file through UVLE.

8x8 Handwritten Digits Visualization and Classification

For this problem, we'll use the example given in Lecture 4: 8x8 handwritten digits data. Load the `load_digits` data from sklearn.datasets.

Do the following:

a. **[30 pts]** Normalize the X data using Standard Scaler. Then, project all the X data into 2 dimensions using 6 dimensionality reduction techniques:
      i. Local Linear Embedding (n_neighbors = 200, random_state = 0)
      ii. t-SNE (perplexity = 50, random_state = 0)
      iii. Isomap (n_neighbors = 200)
      iv. Laplacian Eigenmap (n_neighbors = 200)
      v. Kernel PCA (kernel = 'rbf', gamma = 0.01)
      vi. PCA

    The points should then be colored according to the digit labels, y.

    Which of the methods produced clear clusters of data points?

b. **[10 pts]** For both KPCA and PCA in the previous item, report the CPV plot and determine the number of principal components needed to retain 95% cumulative percent variance.

c. In this item, we will perform classification with and without dimensionality reduction. First, split the data into 70% training and 30% testing, *stratified* according to class label. Compare the results between the following methods:

      i. **[20 pts]** Make a pipeline using StandardScaler, Kernel PCA (kernel = 'sigmoid', n_components=40), and SVC (default hyper-parameters). Fit the pipeline on the training set, then report the accuracy and F1-score on the test set.

      ii. **[20 pts]** Make a pipeline using StandardScaler, Linear Discriminant Analysis (n_components=9), and SVC (default hyper-parameters). Fit the pipeline on the training set, then report the accuracy and F1-score on the test set. Can we extract more than 9 components from LDA? Why or why not?

      iii. **[20 pts]** Make a pipeline using StandardScaler and SVC (default hyper-parameters) alone. Fit the pipeline on the training set, then report the accuracy and F1-score on the test set. This is the case without any dimensionality reduction.

    Which of the three methods had the best test performance for classification? Explain.

<div align="center">END OF EXERCISE</div>