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]** Using PCA, visualize the projection of the data set onto the first 2 principal components. In the plot, color the points based on their class label. Finally, generate the explained_variance_ratio plot (or CPV plot) of the data. What is the CPV at 2 PCs?

b. **[50 pts]** From the 2-D result in item (a), split the data into 70% Training and 30% Testing with stratification. Use random_state=0 in the train_test_split function. Train an SVM with your own hyper-parameter tuning scheme. Report the classification accuracy and confusion matrix for both the Training and Test sets based on the best tuned model. **Remember:** The X data should only have 2 columns (features) rather than 64, upon feeding it to SVM.

c. **[20 pts]** Instead of PCA + SVM, perform LDA with 2 components on the original 64-feature data set. Split the data into 70%-30% training/testing with stratification and random_state=0 (similar to item b). Plot the projected data in 2-D space with colors corresponding to their class labels; use a circle marker for training data, then x marker for test data. Finally, report the classification accuracy and confusion matrix for both Training and Test sets based on the LDA predictions.

Based on your results, which of the two methods above (PCA+SVM vs. LDA) is more preferrable for the 8x8 Handwritten Digits classification? Use various aspects for comparison such as computational effort, human effort, model accuracy, interpretability, etc.

END OF EXERCISE