**Final Project**

**Project description:**

1. Using Python and Numpy, write a class named KMeans with necessary methods, which implements the K-means method for clustering.
2. Using Python and Numpy, write a class named SoftKMeans with necessary methods, which implements the soft K-means method for clustering.
3. Consider the wheat seed data set as given in seeds\_dataset.txt. This dataset contains 210 data samples. Each sample has seven input features and one output label, which is described in the following table.



Use K-means and soft K-means for clustering analysis with the above data set. As shown in the dataset, the wheat has three varieties, namely, Kama, Rosa, and Canadian, labeled 1, 2, and 3, respectively. So firstly, we set K=3 for clustering. But it should be noted that we do not need these label information for K-means and soft K-means method. We only use them to measure the performance of clustering. Compare the performance of K-means and soft K-means for this experiment.

1. Consider the above dataset again and try to set K=10. Then compare K-means and soft K-means and demonstrate. Moreover, modify the two algorithms by adding non-local split-and-merge moves, and then run the modified algorithms for this experiment with K=10. Explain what you observed from the experiment.

1. Using Python and Numpy, write a class named PCA with necessary methods, which implements a standard PCA method.
2. Using Python and Numpy, write a class named LinearAutoEncoder with necessary methods, which implements a linear autoencoder.
3. Find an image and use the PCA method as well as the linear autoencoder method to find the basis of principal components (you could try with different dimensions). Moreover, compare the reconstructions. Finally, use soft K-means to get clustering results with K to be the same values as the dimension of principal components in PCA.

**Requirement:**

1. You should complete your project by yourself only, and then hand in your code as well as a project report before the deadline.
2. The final project takes 20 marks for the course (20%).
3. The deadline to submit your project report and code packages is 23:59PM of Dec. 24, 2023. It is a firm deadline (Late submission will receive 0 mark).
4. When completing your course project, you are required to write a project report together with the codes for the project. Base on the project report and the code package, the project will be marked.
5. The project report should be written in English.
6. The project report should be presented in the IEEE conference paper style and suggest to use LaTex if possible. Refer to the following link

https://www.ieee.org/conferences/publishing/templates.html

for the LaTex Template (a LaTex template package is also included in the zipped file), or you can work in Overleaf (an online LaTex editor). The project report should contain the project title, author, abstract, keywords, I. Introduction, II. Problem formulation, III. Method and algorithms, IV. Experiment results and analysis, V. Conclusion and future problems, and possibly References.

1. Hand in a complete code package including the data set, code files with detailed description of dependencies, etc., so that the code can be checked and run on another computer without any problem.
2. The project and the codes should not be copied from others. Once it is noticed that the hand-in is copied from others including your classmates or online available work, you will receive 0 mark.