Week 8 – Homework

Requirements:

1. **Explain** how Gaussian Mixture Models are used to cluster unlabeled data.
2. **Implement** Gaussian Mixture Models Clustering Algorithm in Python. Employ your implementation on the given dataset in Canvas (The dataset is given in the file named "GMM_EM_data_for_clustering.csv".). Compare the results by changing the initial parameters (you can use random initialization and then repeat your implementation) and make discussion.
3. **Evaluate** the results by comparing your implementation with the well-developed toolbox-- scikit-learn and make discussion.

Description:

1. **Source code** of your implementation of GMM Clustering Algorithm.
2. **A report** that evaluates your implementation and the toolbox. The report should **NOT EXCEED** 2 pages. The report is recommended to include Title, Author(s), Methods, Experimental Settings, and Results & Discussion. It can be written in English or Chinese. Please use the IEEE template for journal/transactions to format the report. In general, the report is required to be formatted using double columns, a font size of 10pt and single line space.

i) Title.
ii) Author(s): This homework can be finished by one individual OR a team of two. If you choose to team with others, please specify your respective contributions in the report.
iii) Methods: Algorithmic description of GMM Clustering.
iv) Experimental settings: Your platform/environment and the initial parameters you set.
v) Results & Discussion.

Submission:

Please submit the source code and report via Canvas **NO LATER THAN May 5, 2021** (11:59 PM UTC +8). Pack the report and source code into a zip or rar file with filename "your name(s)_your ID(s)".

Important Notice:
Please do not duplicate complete sentence(s)/paragraph(s), figures, and tables from publicly available sources without citation, which would cause a potential risk of plagiarism. If it is necessary to do so, please mark the duplication and cite the source.