Week 9 – Homework

Requirements:

1. **Implement** PCA in python and apply it to the MNIST dataset (The dataset is given in Canvas named "mnist-original.mat").
2. **Develop** an algorithm to implement PCA in a more efficient manner to handle the case when we have *the dimensionality greater than the size of the dataset*.
3. **Evaluate** the results by comparing the two implementations and make discussion.
4. **Compare** your implementations with the well-developed toolbox-- scikit-learn and make discussion.

Description:

1. **Source code** of your implementation of PCA.
2. **A report** that evaluates your two implementations and the toolbox. The report should **NOT EXCEED** 3 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 the two versions of PCA. For the case when we have the dimensionality greater than the size of the dataset, you need to give the derivation.

iv) Experimental settings: The number of samples you used in the two implementations, the platform/environment, etc.

v) Results & Discussion.

Submission:

Please submit the source code and report via Canvas **NO LATER THAN May 10, 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.