From: **ZHANG Fa** fa.zhang@connect.ust.hk
Subject: CSIC 5011: Project 1 Review
Date: 19 April 2023 at 9:10 PM
To: datascience.hw@gmail.com

Dear Prof. Yao,

Hi, my name is Fa Zhang (20823094) and below are my comments of five randomly chosen reports.

1.
02 Image Inpainting with PCA.
Summary of the report:
This report uses components learned from PCA to construct the dictionary to do inpainting for hand-written digits dataset.
Describe the strengths and weaknesses of the report:
This method can restore semantic information for moderate damage and the report is well visualized.
This report didn't compare with other methods.
Evaluations: 5 5 5 2

2.
05 Visualization and Dimensionality Reduction Techniques for US Crime Data.
Summary of the report:
This report uses different methods to perform visualization and dimensionality reduction for US crime data.
Describe the strengths and weaknesses of the report:
This report tests many techniques, PCA, Sparse PCA, MDS, ISOMAP and LLE.
More analysis can be done to show the difference between these five methods and why is that.
Evaluations: 4 3 3 2

3.
11 Visualization and Forecasting on Finance Data.
Summary of the report:
This report uses different methods to perform visualization and dimensionality reduction for SNP'500 dataset.
Describe the strengths and weaknesses of the report:
This report tests many techniques, PCA, SPCA, MDS, ISOMAP and t-SNE.
Instead of using MSE, use the classification accuracy to compare the performance of different models may be better.
Evaluations: 5 4 4 2

4.
14 Topic Modeling For NIPS Words.
Summary of the report:
This report uses latent Dirichlet allocation (LDA) to do dimension reduction and learn topic distribution of NIPS words dataset.
Describe the strengths and weaknesses of the report:
The LDA model is very powerful and suitable for this task.
PCA can be applied to also learn the latent topic and then be compared with LDA model.
Evaluations: 5 5 5 2

5.
17 Exploration of PCA family for handwritten digit classification.
Summary of the report:
This report investigates the performance of different kinds of PCA on hand-written digits dataset.
Describe the strengths and weaknesses of the report:

Used methods are well introduced and visualized.

The result show the original RF achieves the highest accuracy, which doesn't show the benefit of dimension reduction.

Evaluations: 5 3 3 2

Best,
Fa