From: **TENG Fei** fteng@connect.ust.hk
Subject: CSIC 5011: Project 1 Review.
Date: 21 April 2023 at 11:36 AM
To: datascience.hw@gmail.com

Qiqi Zhou

Summary:

This project deals with inpainting problem via machine learning approaches with the help of dimension reduction techniques. The experiment results illustrate that the proposed method can help to recover the semantic information to some extent.

Strengths:

The project is done in a solid manner. The author firstly presents the damage patterns. Then depicts experiments results with diagrams with detailed analysis given.

The author tuned hyperparameters in a reasonable way instead of exhaustive search.

The visualization result are clear and easy to understand.

Weakness:

The situations of damage pattern only covers row, column and cross are limited. The performance is limited when dealing with cross. Will the performance drops severely when other more complex shows?

The experiment results are presented by visualizable diagrams. But no evaluation metrics are given.

Scores:

Writing: 5.

Technology: 5. Though lacks of evaluation metrics, the visualization is clear enough.

Overall: 5.

Confidence: 3.

Guangzheng ZHONG

Summary:

This project proposes a dimension reduction method for ordinal regression dimension reduction, aiming to solve the distance problem in multi-class classification problem. Specifically, the author propose a new distance function that considering the class information in order to apply dimension reduction algorithms on supervised classification task.

Strengths:

The solid methods illustrated with clear visualization results.

The overall writing is detailed and well-organized.

The experiments are solid that tests the distance function's performance with 3 dimension reduction techniques.

Weakness:

The situations of damage pattern only covers row, column and cross are limited. The performance is limited when dealing with cross. Will the performance drops severely when other more complex shows?

The experiment results are presented by visualizable diagrams. But no evaluation metrics are given.

Scores:

Writing: 5. The writing is well-organized with comprehensive diagrams.

Technology: 5. The overall technology are solid and with complete tests.

Overall: 5.

Confidence: 3.

Ruochen MA; Jihong TANG; Yuyan RUAN; Zhi HUANG

Summary:

This project is to test the performance for seven different dimension reduction techniques via ancestry prediction on SNP dataset. They proved that random projection outperforms other six dimension reduction approaches.

Strengths:

It depicts the visualization result of dimension reduction. Then applying line chart to depict the performance of dimension reduction approaches, PCA, MDS, t-SNE, ISOMAP, LLE, UMAP, robust

PCA and random projection in a solid way.

Weaknesses:

It is quite confusing that why divide people by regions. Will the experiment results vary if selecting other dividing criteria?

Scores:

Writing: 4. Lack of discussion of reasons why random projection outperforms other methods.

Technology: 4. No explanation of division criteria.

Overall: 4.

Confidence: 3.

Fa ZHANG and Ruizhe XIA

Summary:

This project is to test the effectivenss for shrinkage methods to improve the performance of Mean-variance optimization portfolio problem. It proves the existing findings that the performances of GMVP, CM1, and CM2 are comparable to the MVP, though the inverse will still lead to the performance drop.

Strengths:

It clearly reports the metrics for several covariance matrix shrinkage and precision matrix shrinkage on synthetic dataset and finance dataset respectively in a solid manner. All choices parameters.

The overall writing is well-organized and detailed with background information provided.

Weaknesses:

Some choices of hyperparameters are not given. Causing some difficulties for readers. What will happen if changing them?

Scores:

Writing: 4. The writing is detailed and well-organized with clear formula and background presented. But some hyperparameters choices are confusing.

Technology: 5. The overall technology are solid and with reproducible tests.

Overall: 4.5.

Confidence: 2.

Lue SHEN

Summary:

This project is to apply dimension reduction techniques to analyze the single nucleotide polymorphism data. By the obtained importance of features, duduce the reason for human migration. Specifically, the author compares the performance for different dimension reduction techniques. Finding that the clustering results are consistent to African origin hypothesis.

Strengths:

It clearly reports the results for different dimension reduction methods result of human clustering on SNP dataset by multiple-form diagrams and charts.

The overall writing is well-organized and provided detailed description and formula for each dimension reduction method.

The experiments provide interesting insights for analyzing human migration problem.

Weaknesses:

It seems that the sample size will affect the experiment result. But there exists few experiments to test relative large sample size.

Scores:

Writing: 5. The writing is detailed and well-organized with clear formula and background presented.

Technology: 4. Lack enough experiments for feature sample size.

Overall: 4.5.

Confidence: 3.

review.pdf