
Reviews of Mid-term Mini-projects MATH-5473

Hangyu Lin
Department of Mathematics
hlinbh@connect.ust.hk

The section title Review 1: 4552 means the final evaluation scores are 4, 5, 5, 2 for Review 1.

1 Review 1: 4552

03. YAN, Ningyu; YAN, Bokai; LAI, Yanming
The statistic explanation of the US crime data.

1.1 Summary

They analyzed the US crime data by different methods like PCA, IsoMap, UMAP. Specifically, they first use PCA, Isomap UMAP to reduce the dimension of the original data and then use linear regression to fit the data. Based on the experiments, they found some meaningful factors that affect the crime rate.

1.2 Strengths and Weaknesses

Strengths

- The plots and pictures are easy to understand and intuitive to navigate.
- The evaluation and validation for different methods are consistent and significant.

Weaknesses

- More detailed discussion about the results will be better, like section 3.3, if there are some explanations related to the crime it will be more clear.

1.3 Evaluation

Evaluation on Clarity and quality of writing: 4

Evaluation on Technical Quality: 5

Overall rating: 5

Confidence on your assessment: 2

2 Review 2: 5552

09. HUANG, Zhanmiao; Wencan XIA; Yuanhui LUO. Explore Geographical Variation of Human Gene with SNPs Data.

2.1 Summary

In this project, they explored the geographical variation of human genes with the SNPs data. They applied PCA, and MDS to visualize the data and then use the random forest and extra trees to classify the data. They give some results related to human genes and regions based on the results of these methods.

2.2 Strengths and Weaknesses

Strengths

- They conduct a thorough analysis with PCA and MDS combined with the statistical learning method, and the results are significant.
- The case study are reasonable and give some reliable conclusions based on the results.

Weaknesses

- Some quantitative results can be added to further improve the analysis.

2.3 Evaluation

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality: 5

Overall rating: 5

Confidence on your assessment:2

3 Review 3: 4442

11. HAO, Yifan; Yueying Hu; Yakun Li; Yonglin Liu.
Visualization and Forecasting on Finance Data.

3.1 Summary

The report used PCA, MDS, ISOMAP, t-SNE to visualize SNP 500 data and then try to forecast the results by using PCA, SPCA, target-PCA.

3.2 Strengths and Weaknesses

Strengths

- The processing and visualization for time series are clear.

Weaknesses

- For each method, more explanation for the results is needed to understand what it means.

3.3 Evaluation

Evaluation on Clarity and quality of writing: 4

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence on your assessment:2

4 Review 4: 4442

12. DUNDA, Gerry Windiarto Mohamad. Statistical Analysis on Authors and Word Trend of NIPS Papers from 1987 to 2017.

4.1 Summary

The author studies the statistical summary of authors and word trends for NIPS papers from 1987 to 2017. In the report, they found a weak positive relationship between community size and the average number of publications and clustered papers using MDS and work mover's distance.

4.2 Strengths and Weaknesses

Strengths

- The research question and methods used are clear.

Weaknesses

- Maybe more visualization-based methods like PCA can be used to visualize the data.

4.3 Evaluation

Evaluation on Clarity and quality of writing: 4

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence on your assessment:2

5 Review 5: 4442

13. LI Haobo, CHEN Zixin, TENG fei, SHENG Rui. NIPS Conference Papers 1987-2015 Data Set.

5.1 Summary

They analyze the NIPS conference papers data and try to visualize the word distribution during 1987-2014. By leveraging BERT, HMM, and t-SNE, they extract time series-based analysis and frequent words prediction.

5.2 Strengths and Weaknesses

- The plots and pictures are easy to understand and intuitive to navigate.
- The model structure is reasonable.

Weaknesses

- All the results are based on a well-designed model, maybe some comparison with the traditional data reduction method can be conducted to make the whole results more complete.

5.3 Evaluation

Evaluation on Clarity and quality of writing: 4

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence on your assessment:2