CSIC 5011 mini project review

WANG Zhiwei, 20822753

April 21, 2023

02. ZHOU Qiqi. Image Inpainting with PCA.

This project explored the image inpainting problem, which is a common, important, and fundamental task in computer vision and image processing. It involves filling in missing or damaged parts of an image, which can occur for various reasons, such as noise, occlusion, or physical damage. The author used components learned from PCA to serve as the dictionary and performed inpainting well in the handwriting digits dataset.

The report is clear and logical, and the results are very impressive and well-organized. However, the experiments part is slightly insufficient. More evidence is expected to be provided to prove the concept and fully convince the reviewers, including the compared methods.

- Clarity and quality of writing: 5.
- Technical Quality: 4.
- Overall rating: 4.
- Confidence on my assessment: 3.

05. XU Yingxue; Jiaxin ZHUANG; Fengtao ZHOU. Visualization and Dimensionality Reduction Techniques for US Crime Data.

In this project, the authors aimed to gain a deeper understanding and insights into the US crime dataset. They applied Principal Component Analysis (PCA) to reduce the variable dimensions from 24 to 2/3 by selecting the top two/three eigenvalues and eigenvectors. The results showed that the crime number is influenced by all the factors, such as the number of sworn police officers employed by the city, the number of civilian police employees, and others.

The report was well written but a bit cumbersome and wordy, making it difficult to capture the key points. In addition, the total workload is slightly insufficient, and of course, the results are also. The subsequent analysis also lacked a deeper understanding of this dataset.

- Clarity and quality of writing: 2.
- Technical Quality: 3.

- Overall rating: 2.
- Confidence on my assessment: 3.

09. HUANG, Zhanmiao; Wencan XIA; Yuanhui LUO. Ancestry Prediction via Dimensionality Reduction Techniques on SNPs Data.

In this project, the authors conduct PCA, MDS, and random projections on SNPs dataset to explore the genetic variation with geographic variations. The results indicate that both PCA and MDS can separate people from different regions based on essential principal components of SNPs, which is more efficient than the random selection of SNPs.

The report is written clearly. The results and figures are well organized and analyzed reasonably.

Here is my question about the statistical methods part. The SNPs dataset is a "wide" dataset, i.e., few samples but many features (in the case here, 650,000 SNPs and 1,064 people), and moreover, the noise is extremely high. However, tree-based methods, such as random forest, are not stable when handling this kind of data. The linkage disequilibrium (LD), i.e., the correlation between the features, will also affect the results. So I'm concerned about the overfitting problem and reliability in this part. Whether the fitted model and the results of the analysis are reliable. Whether the results will be consistent when using different datasets?

- Clarity and quality of writing: 4.
- Technical Quality: 4.
- Overall rating: 4.
- Confidence on my assessment: 3.

10. ZHANG Fa and XIA Ruizhe. Application of Shrinkage in Portfolio Optimization.

The report raised an important question-covariance matrix estimation and precision matrix, which has attracted a great deal of attention in both the finance and optimization field. Not only does this problem imply deep financial background knowledge, but it is also well-defined mathematically. With the classical statistics principle, "bias-variance trade-off", the authors used the shrinkage techniques and achieved very impressive performance.

The writing is clear and easy to understand, and the results, figures, and tables are well organized. However, the experiments part is slightly insufficient. More evidence is expected to be provided to prove the concept and fully convince the reviewers.

- Clarity and quality of writing: 5.
- Technical Quality: 5.
- Overall rating: 5.
- Confidence on my assessment: 3.

16. Chris HC Nguyen and James M Shihua. Robust PCA for Moving Object Detection in Video.

In this project, the authors demonstrated the Robust PCA (RPCA) algorithm for the detection and removal of moving objects in a video. The algorithm converges very fast and is significantly more efficient than traditional methods like the alternating direction method of multipliers (ADMM). From the output foreground and background components, it's clear that the algorithm successfully extracted the moving objects (people) and stationary objects (floor, shelves) with a satisfactory level of accuracy.

The GitHub repository is well-organized and looks professional. The report is short but logical and clear, although there are some grammatical errors. But at the same time, the content is not rich enough, and more analysis of the results is expected.

- Clarity and quality of writing: 2.
- Technical Quality: 4.
- Overall rating: 3.
- Confidence on my assessment: 3.