MATH 6380Q Project 1 Peer Review

05. SUN, Jing and LUO, Shuang. Principal Component Analysis of Crime Data in USA.

[ Reviewer 1 ]

1. Principle Component Analysis of Crime Data in USA

Summary:

In this project, PCA was employed to analyze the crime rate of seven different crime types of 59 American cities in 1985. It was found that the crime rate of larceny is the highest among all types of crimes. Moreover, certain crimes were found to have a positive correlation among themselves.

Strength of the project:

Clear logic flow and well-explained for each step.

|  |  |
| --- | --- |
| Evaluation on Clarity and quality of writing (1-5): | 5 |
| Evaluation on Technical Quality (1-5):  | 4 |
| Overall rating:  | 4 |
| Confidence on your assessment: | 2 |

[ Reviewer 2 ]

In poster five, “Principal Component Analysis of Crime Data in USA”, author used PCA method to analyze the crime data which consist of the crime rate of seven different crime types of 59 American cities in 1985.Athor first raised the background of the question and then used PCA to get the results. Results show that the crime rate of larceny is the highest among all the types of crimes. From the top three principal component embedding, it is found that cities including Newark, Portland, Miami and Detroit are the most common outliers. In addition, crime types including assault and robbery, auto theft and murder, auto theft and robbery, larceny and burglary show a positive correlation.

The strength of this poster is the poster is well organized and there is an abstract which is very clear. The results are convinced and technical sound. The research questions have social significance and the results obtained are instructive.

The weakness of this poster is not well enough.

Evaluation on Clarity and quality of writing: 4. The report is clearly written and the poster is well organized. And there is no suggestion to improve the clarity of the paper.

Evaluation on Technical Quality: 4.

Overall rating: 5.

Confidence on my assessment: 3.

[ Reviewer 3 ]

05. SUN, Jing and LUO, Shuang. Principal Component Analysis of Crime Data in USA.

The authors used the PCA to explore the variance of the crime rate in different cities of the States. It was found that larceny had the highest crime rate. With top 3 component embedding, Newark, Portland, Miami and Detroit are the most outliers.

Apart from PCA, the authors also did some statistics analysis on the crime rate data.

One thing may be confusing is that what’s the meaning of the outliers in some plots.

For the writing, I suggest that they should use the past tense apart in addition to actual facts. Also, add more facts and examples from the news report or other investigations to explain the results. Since you are not doing any collecting data or simulation work, do not include such description in your acknowledge part.

The results are technically sound. For explaining the variance, they’d better plot the explained variance ratio versus PCA component rather than the eigenvalue since it is not intuitionistic. I checked the codes and they are fine. Like many other reports, they did not refer to the weakness of these methods. The reference is complete.

The overall rating is 3 and my confidence is 2.

[ Reviewer 4 ]

This report gives an exhaustive analysis of the crime data in USA. It has many figures and tables to support its conclusions, and the reasoning is sound.

Evaluation on Clarity and quality of writing: 5

Evaluation on Technical Quality:4

Overall rating:4

Confidence on your assessment:2

[ Reviewer 5 ]

Summary: In this project, Principal component analysis (PCA) is employed to explore the variance of the crime rate existed in different cities by using covariance matrix decomposition method.

Strengths: Reasonable explanation for the first 3 eigenvalues.

Weakness: More explanation on Fig 6,7 will be better.

Evaluation on Clarity and quality of writing: 4.

Evaluation on Technical Quality: 3.

Overall rating: 3

Confidence on your assessment: 2

[ Reviewer 6 ]

4.1 Summary

For Sun Jing and Luo Shuang’s work, they applied Principle Component Analysis on the crime dataset. They investigated the crime rate for different crime type, and use principle component analysis to determine the city outliers, and analyze the correlation between each crime types.

4.2 Strength and Weakness

Luo and Sun constructed a complete work for the Principle Component analysis on the crime data. In the content of their report, the briefly introduction of crime and PCA are included. The one weakness I can tell is the lack of parallel analysis which can be applied.

4.3 Score

4.3.1 Clarity and Quality of Writing

I appreciate their effort on explaining all the important concept in their report, even unprofessional people can understand their report. The structure of the report is clear apparently. I will give them 5/5 on this aspect.

4.3.2 Technical Quality Reasonable using of different method. Their work investigate a topic by using different methods in a feasible way. The content can be more integrated if they include parallel analysis. If they can mention about future work it would be better. I will give them 4/5 on this aspect.

4.3.3 Overall

The overall score for this report is 4.5/5.

[ Reviewer 8 ]

Summary of the report.

Use PCA to analyze the crime data.

• Describe the strengths of the report.

Clearly and detailedly written. Enough analysis of the results.

• Describe the weaknesses of the report.

The methods (PCA/visualization and correlation analysis) might be a little bit simple. (I think)

• Evaluation on Clarity and quality of writing (1-5): Is the report clearly written? Is there a good use of examples and ﬁgures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the clarity of the paper, and provide details of typos.

5

• Evaluation on Technical Quality (1-5): Are the results technically sound? Are there obvious ﬂaws in the reasoning? Are claims well-supported by theoretical analysis or experimental results? Are the experiments well thought out and convincing? Will it be possible for other researchers to replicate these results? Is the evaluation appropriate? Did the authors clearly assess both the strengths and weaknesses of their approach? Are relevant papers cited, discussed, and compared to the presented work?

3

• Overall rating: (5- My vote as the best-report. 4- A good report. 3- An average one. 2below average. 1- a poorly written one).

4

• Conﬁdence on your assessment (1-3) (3- I have carefully read the paper and checked the results, 2- I just browse the paper without checking the details, 1- My assessment can be wrong)

2

[ Reviewer 9 ]

[ Reviewer 10 ]

*Summary:*

Use PCA to analyze the crime in US.

*Strength:*

Tell a sound story with application of PCA, and get meaningful result from it. The background and methodology describe clearly.

*Weakness:*

The analysis of PCA result could be better organized maybe.

*Evaluation on Clarity and quality of writing (1-5): 3*

(format) The space left before paragraph and subtitles seems not necessary, and some paragraph has space before while some do not.

*Evaluation on Technical Quality (1-5): 4*

The crime data seems complicated enough to analyze with PCA or correlation matrix, to draw meaningful result.

The method is standard, and they have a good interpretation of result.

*Overall rating (1-5): 4*

*Conﬁdence on your assessment (1-3): 2*

[ Reviewer 11 ]

Summary: This report analyses the crime data in American cities. They found that there exists some common outliers and some crime types have positive correlation.

Strengths: The experiments and results are very clear.

Weaknesses:

Writing: (5) The report is well-written and well-organized.

Technical Quality: (4) The claims are well-sported by the experiments. I have a small question that the covariance matrix is calculated using raw data or first three principal components?

Overall rating: (4)

Confidence on my assessment: (2)

[ Reviewer 12 ]

**Summary:** This project uses PCA to analyze the crime data of different types in different cities in a single year. By analyzing the first, second and third principal component scores and corresponding coefficients, it draws the conclusion about which type of crimes contribute the most variance. It also explores the correlation between different types of crimes with covariance matrix.

**Strengths:** This project uses various types of figures and tables to show its results. The visualization part is well done. It analyzes the scores and coefficients carefully and tries to obtain as much information as possible from limited data.

**Weaknesses:** Only PCA is utilized in this project which makes the content not so rich. Besides, some of the figures are hard to understand and adding more detailed legends may help to improve.

**Evaluation on Clarity and quality of writing:** 4

**Evaluation on Technical Quality:** 4

**Overall rating:** 4

**Confidence on your assessment:** 2