MATH5473 MINI REVIEW

CHEN, Zixin 20582896

April 2023

1 Image Inpainting with PCA

Author: Qiqi ZHOU

Summary

This project utilizing PCA to construct a dictionary for inpainting images on the hand-written digits dataset, which is one of the crucial steps in the pipeline without Deep Learning. The author first applied PCA to find the suitable number of k for the selection of top-k important components. Then, the author use k-means clustering to get a dictionary which can depict the distribution of images. Finally, the author did a two-step transformation, converting the damaged image into the base space to find the most similar undamaged data and then converting the representation back to inpaint the image.

Strength of the report

This report is well-written with a clear description in methodologies, steps, experiment details, and interpretations. The reader can easily follow the author's workflow to see how this significant issue is solved. Moreover, the problem description and background introduction is a shining point, in which the author uses precise language to draw a whole background picture for the problem he raised.

Weakness of the report

One potential weakness of this report might be the type-setting of the figures. Specifically, the caption for Figure 5-7 is slightly unfriendly to readers. I understand it is due to the limitation of the page size, but personally I believe a clear figure can win a detailed description. Thus, I would recommend the author to reduce some texts in order to provide enough room for your figures, especially the ones to show your results.

Evaluation of Clarity and quality of writing: 4.5

The use of examples and figures are good and well-organized. One tiny problem has been mentioned in the weakness.

Evaluation of Technical Quality: 5

The reproducible code are provided. The results are technically sound. The author's claims are well-supported by both theoretical analysis and experimental results. The references are also clearly presented.

Overall rating: 5

Confidence on my assessment: 2.5

2 Ancestry Prediction via Dimensionality Reduction Techniques on SNPs Data

Author: Ruochen MA, Jihong TANG, Yuyan RUAN , Zhi HUANG

Summary

This project utilizes several popular dimensional reduction techniques on SNPs dataset to make ancestry predictions. The authors first tried different methods to separating people frin 7 different regions, then chose method t-SNE to do further prediction on an additional 1000 Genomes Project dataset.

Strength of the report

The report is consisted of plenty of experiments and all the results are clearly provided. The authors tried and compared several alternative methods at each step of their workflow and showed detailed illustrations. The experiments are solid with detailed explanations. The choice of figures helped a lot to support their statements.

Weakness of the report

As the authors utilized many alternative methods and in Part 2, RandProj performed best under two criteria, I am confused about the choice of t-SNE in Part 3. The potential reason might be t-SNE is better at visualization clusters, but some reasons are required to make the logic flow more coherent.

Evaluation of Clarity and quality of writing: 4

The use of examples and figures is good and well-organized. The logic transition from part 2 to part 3 might be able to improve.

Evaluation of Technical Quality: 5

The reproducible code are provided. The results are technically sound. The authors used plenty of methods in each steps of their project, which is very solid

Overall rating: 4.5

Confidence on my assessment: 2.5

3 Visualization and Dimensionality Reduction Techniques for US Crime Data

Author: Yingxue XU, Jiaxin ZHUANG, Fengtao ZHOU

Summary

The authors utilized PCA and several manifold learning methods and meticulous approaches to explore the US Crime Dataset. The authors successfully reduce 2/3 of the dimensions of the dataset and find out the relations between each variable and the crime number. The authors finally draw the conclusion that PCA has good performance in terms of data reduction and visualization.

Strength of the report

The experimental results are detailed and clear. The authors provided a lot of explanations for each of the experiment figures. Also, it's good to know the data preprocessing process the authors used, which could help readers to understand the source of 24 features.

Weakness of the report. For the experimental results, beyond detailed explanation of each results, maybe some more comparison could help to further illustrate the strength and weakness of each methods. Also, I am personally curious about the result of LLE in figure 1(e), as we could only see 3 dots on the result. As for the conclusion, personal opinion is: a comparison between different methods might be better than simply claim that PCA is useful for dimension reduction, as it is a somehow well-known conclusion.

Evaluation of Clarity and quality of writing: 4

The use of examples and figures is good and well-organized. The report is

clear and easy to follow.

Evaluation of Technical Quality: 3.5

The reproducible code are provided. The authors did several alternative experiments to try different methods. Personally, I think some more advanced techniques or some combination of different methods might be also useful.

Overall rating: 4

Confidence on my assessment: 2.5

4   Robust PCA for Moving Object Detection in Video

Author: Chris HC Nguyen and James M Shihua

Summary

The authors utilize RPCA to identify and eliminate the moving objects from the background of the video. Detailed algorithms and figures are provided to show their results after numbers of iterations of the algorithm running. Finally, they lead to the conclusion that RPCA is useful for extract the foreground and background components from a video.

Strength of the report

The figures are very helpful to show their results. The authors also provide a detailed introduction about the data and the methods they used. The logic of the poster is clear and easy to follow. And the conclusion also mentions some limitations which is good to learn from.

Weakness of the report

Personally, I don't think provide too detailed algorithm formulas is the best way of utilizing the space of a poster. As the authors already provide their source code, some comprehensive explanation of their methods might be a better choice. On the other hand, maybe providing more than one case in the iteration result is more suitable for the content, as the aim of this project is to identify and eliminate moving objects from a video, rather than a picture.

Evaluation of Clarity and quality of writing: 4

The use of examples and figures is good and well-organized. The poster is clear and easy to follow. The type-setting is good and the report looks elegant.

Evaluation of Technical Quality: 4

The reproducible code is provided. Maybe some alternative methods and their comparison could also be included. The technical details are shown clearly.

Overall rating: 4

Confidence on my assessment: 2.5

5 EXPLORING THE EFFECTIVENESS OF PCA ON HANDWRITTEN DIGIT DATASET

Author: JIA Guangnan

Summary

The author explore the effectiveness of PCA on the handwritten digit dataset. Meanwhile, the author also explore the effect of using K-means clustering and logistic regression on handwritten digit recognition. The results show that PCA is an effective technique for reducing the dimensionality of the dataset while preserving the essential information in the data, as the images from reconstruction is quite successful compared to the original images. Finally, a comparison between K-means and logistic regression is conducted.

Strength of the report

The author provided a detailed background introduction with clear problem setting, objectives, methods to be applied and data preprocessing. For the experimental part, the use of figures help the readers to understand the results. It's good to visualize different PCA components and provide brief descriptions of some of them, which enable the readers to see the effectiveness of PCA.

Weakness of the report

Personally thinking, some more advanced methods might be discussed to improve the quality of the results. The reasons for choosing K-means and logistic regression as a comparison might be added. Also, the author might consider including a reference list.

Evaluation of Clarity and quality of writing: 4

The use of examples and figures is good and well-organized. The author explained the details very clearly. The reasons for the choice of each algorithm might be better to be included.

Evaluation of Technical Quality: 4

The reproducible code is provided. Maybe some alternative methods and their comparison could also be included. The technical details are shown clearly.

Overall rating: 4

Confidence on my assessment: 2.5