MATH 6380Q Project 1 Peer Review

04. SHEN, Xinwei and YANG, Yunfei. Dimension reduction methods to improve image classification.

[ Reviewer 1 ]

1. Dimension reduction methods to improve image classification

Summary:

The project is divided into 3 parts. The first part investigates the effect of performing PCA as data preprocessing on the subsequent classification efficiency. It was found that while increasing the sample size to certain level, the top k-PC SVM has even better performance than the SVM model trained with raw data.

The second part of the study compares the accuracy of the regularized estimators with traditional MLE, and they found that the estimators with L1 penalty out-performs the others, indicating the sparse underlying structure of the image dataset. In the last part, they found that combining PCA and LASSO could yield even better accuracy than LASSO alone.

Strength of the project:

A very comprehensive theoretical case-study using MNIST dataset. The part where combing PCA and LASSO could yield an even better result is particularly inspiring.

Weakness of the project:

I found the part of PC.Adjust quite confusing, as far as I could understand, the difference between PC.Adjust and PC.Regression is that the former one only contains top-5 PCs while the latter contains top-50 PCs, it is suggested that the naming of the two candidates could be improved to avoid the confusion. And in this case, as a fair comparison, one more candidate could be added to the experiment, which is the LASSO combined with top-50 PCs.

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

[ Reviewer 2 ]

In poster four, “Dimension reduction methods to improve image classification”, author designed three different experiments and compared many dimension reduction methods such as principal component analysis (PCA), as well as regularized estimators including Lasso, Ridge and Elastic Net. They explore the power of dimension reduction techniques to improve image classification accuracy more efficiently, and briefly analyzed the reasons. The results are in high dimensional settings, PCA may fail to capture the signals and all the regularized estimators generalized beer than the standard model. The author comprehensively analyzed and compared these methods and I will not repeat that.

The strength of this poster is author used three different experiments to comprehensively compare these methods and the results are convinced. And the conclusion that both geometric dimension reduction techniques and regularized estimators can lead to more eﬀective and eﬃcient on classification performance is very useful.

The weakness of this poster is figures is a bit less so that the result can't be seen very intuitively.

Evaluation on Clarity and quality of writing: 4. The report is clearly written and the poster is well organized. Suggestion is to list figures to make the results more intuitive.

Evaluation on Technical Quality: 4.

Overall rating: 4.

Confidence on my assessment: 3.

[ Reviewer 3 ]

04. SHEN, Xinwei and YANG, Yunfei. Dimension reduction methods to improve image classification.

The authors used dimension reduction techniques PCA/MDS and regularized estimators such as LASSO to do the image classification in a more effective way. They also analyzed the reason behind the good performance.

The authors are able take most methods from the class into consideration and even put forward the theoretical reason behind the results.

One drawback is that in DATA section, more description is needed to make a smooth transition from the MDS result to the following parts.

It is organized very well and follow the structure of a poster. It’s better to add captions these figures.

The results here are well sound and the codes are easy to follow. The three experiments well support the claims in its conclusion part. However, the authors did not touch on the weakness of these methods. The reference part is not well enough since papers on methods used in this poster should all be cited.

It can be rated as 5 and my confidence is 2.

[ Reviewer 4 ]

This report combines different methods such as PCA, Lasso, Ridge and Elastic Net. The information is pretty comprehensive. It would be better if the training time can be listed in the report.

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, the authors explore the power of dimension reduction techniques to improve image classification accuracy more efficiently.

Strengths: Various machine learning techniques are used.

Weakness: N.A.

Evaluation on Clarity and quality of writing: 4.

Evaluation on Technical Quality: 4.

Overall rating: 4

Confidence on your assessment: 2

[ Reviewer 6 ]

3.1 Summary

For Shen Xinwei and Yang Yunfei’s work, they discussed the sample size impact on PCA first, then investigate the accuracy of different regularized estimators. They discover the performance of geometrical data reduction technique and regularized estimators.

3.2 Strength and Weakness

Personally speaking, they have done a excellent job on the comparison part and the analysis section. Each of the experiment is connected to each other, and the construction of the experiment are explained well in the content. The only weakness I can tell is that it will be better to employ one more figure for experiment 3.

3.3 Score

3.3.1 Clarity and Quality of Writing

The structure of the poster is convincing, the analysis part below each experiment is clear as well. They also included the future work they want to do. I will give them 5/5 on this aspect.

3.3.2 Technical Quality

They construct three well-rounded experiment for their investigation. For each experiment, how to get the parameter and how to develop the plot are detaily explained, very good report. I will give them 5/5 on this aspect.

3.3.3 Overall

The overall score for this poster is 5/5.

[ Reviewer 7 ]

* **Summary of this report:** The Hand-written Digits dataset is analyzed by using PCA. It is found that the dimensionality reduction of image features plays a crucial role in image retrieval and classification tasks.
* **Describe the strengths of the report:** The PCA and the regularized estimators are compared. The potential theoretical reasons supporting the results are well analyzed.
* **Describe the weaknesses of the report:** The conclusion in this report is not clearly organized.
* **Evaluation on Clarity and quality of writing (1-5): 5**

No typos are found

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

The technical quality is great.

* **Overal ratings: 5**
* **Confidence on your assessment: 2**

[ Reviewer 9 ]

1. **Summary**

The authors applied dimension reduction methods to improve the I mage classification accuracy. The authors implemented PCA to do the dimension reduction and regularized different estimators.

1. **Strength of the report**

The report is informative and has a good start point and presents the results quite clearly.

Main points are listed with bullet points.

1. **Weakness of the report**

No serious weakness is found.

1. **Evaluation of clarity and quality of writing**

The report is clearly written and well organized. The figures are clearly explained. No obvious problems of the style, grammar, typo, and formatting are examined.

1. **Evaluation on technical quality**

The results are technically sound and no obvious flaws in the reasoning are found. The report provides the methods for analyzing similar dataset. The author compared different estimators and reached the conclusion.

1. **Overall rating**

An overall rating of 5 is given.

1. **Confidence on your assessment**

2- I just browse the paper without checking the details

[ Reviewer 10 ]

*Summary:*

By using different dimension reduction techniques (PCA, Lasso, Ridge and Elastic Net), explore the characteristics of digits dataset.

*Strength:*

Had a deep understanding of the high dimensional classification model and the dimension reduction methods application in model development.

*Weakness:*

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

Maybe uniform the capitalization of PCA in the capitalization form of “PCA”, instead of use lowercase “pca”.

Section of ‘Experiment 3’ is more like a summary and discussion for Experiment 1 and 2.

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

It would be even better to provide the detail of the SVM and MLR model, for replicate of the method.

*Overall rating (1-5): 5*

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

[ Reviewer 12 ]

**Summary:** This project uses dimension reduction methods to preprocess hand-written digits dataset and then uses the reduced features to predict labels and compares their accuracies. It explores the effectiveness of this method and even the potential reasons by trying to do the deconfounding adjustment.

**Strengths:** This project uses PCA as well as other dimension reduction estimators like Lasso. It comprehensively compares the prediction accuracy of these different methods. And furthermore, it tries to remove the hidden confounding factors to get better prediction results. Besides, the poster is neat and clear.

**Weaknesses:** There may be some mistakes about the legends of the last figure where PC-regression (with first 50 PCs) has an accuracy of more than 90% while in experiment 1 it is only about 85%.

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

**Evaluation on Technical Quality:** 4

**Overall rating:** 5

**Confidence on your assessment:** 2