

# Interactive PCA

Mingwei Li

## 1 Introduction

I am going to implement a flavor of PCA [1], that takes users' field knowledge into account in order to plot with more sense. The paper proposed a method that let the user drag data points in an ordinary PCA plot and derive an altered PCA plot from the user's adjustment.

The two challenging parts of this implementation is

1. A probabilistic model of latent variables is involved. It is crucial to get the math correct. I will refer to the PPCA part of Bishop's machine learning textbook [2].
2. It involves building an interactive interface (using d3.js and python flask).

I think the difficulty it is still manageable because once we get the math correct, implementing it is very straight forward. I have written a framework that may be useful[6].

## 2 Relation to Research Work

I am working on visualizing genomic data, specifically trying to build a map of multiple genomes that reflect their genomic affinity. This project is very related to my research because similar work will very likely to help on the interpretability of the genomic plot.

## 3 Expectation

The implementation is expected work similarly on an example given in the paper [1]. i.e. gives similar initial PCA plot (that does not show any clustering) on an education quality data set [3][4], as well as an adjusted plot showing clusters once user drag two points away.

## 4 Comparison and Evaluation

The paper[1] compares the method with some parameter tuning interactions such as iPCA [5], arguing that his method excels because of it is more intuitive to drag points than tuning parameters. I will try to do similar comparisons. However, through this project, I am also willing to evaluate the objectiveness of the resultant plot, because when user feedback is involved, it is hard to judge whether a visualization is showing the truth or a user's belief.

## References

- [1] Endert, Alex, et al. "Observation-level interaction with statistical models for visual analytics." Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011.
- [2] Bishop, Christopher M. "Pattern recognition." Machine Learning 128 (2006): 1-58.
- [3] Guber, D. "Getting what you pay for: The debate over equity in public school expenditures." Journal of Statistics Education 7.2 (1999).
- [4] <https://ww2.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm>
- [5] Jeong, Dong Hyun, et al. "iPCA: An Interactive System for PCABased Visual Analytics." Computer Graphics Forum. Vol. 28. No. 3. Blackwell Publishing Ltd, 2009.
- [6] <https://github.com/tiga1231/snm/tree/master/ppca>