Kernel-Based Dimensionality Reduction Methods on Simulated Datasets

Jonathan Fabish[1], jlfabish@aggies.ncat.edu
Supervisors: Dr. Yishi Wang[2] and Dr. Cuixian Chen[2]

[1] NC A&T State University Department of Applied Mathematics
[2] UNCW Department Mathematics and Statistics

Abstract

High-dimensionality and non-linearity are main challenges of big data sets which can hinder the ability to develop an accurate statistical model. Since every dataset poses unique challenges, there is no one dimensionality reduction (DR) algorithm that does the best job in every case. Therefore, the best approach must be discovered empirically. We are able to engineer synthetic datasets which contain specific geometric challenges. Application of different DR methods to these data, coupled with visualization, is an effective way to gain insight into how to deal with challenges presented by real big datasets.

Objective

In this work, kernel principal component analysis (KPCA), kernel Fisher's discriminant analysis (KFDA), and supervised principal component analysis (SKPCA) were applied to three non-linearly separable, synthesized datasets, using the R statistical programming language. After dimension reduction, the data were projected to a low-dimensional space to gain insight through visual analysis.

Methodology

Linear Dimensionality Reduction Methods

PCA relies on the diagonalization of the covariance matrix of the centered data

$$C = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{x}_j \boldsymbol{x}_j^T.$$

This requires finding the solutions λ and $\mathbf{v}$ to the eigenvalue equation $\lambda\mathbf{v} = C\mathbf{v}$. Finding the eigenvectors that correspond to the largest eigenvalues enables us to find the principal components (PCs), which are the linear combinations of the predictors which explain maximal variation in the dataset. The first PC corresponds to the first eigenvector, each subsequent PC is orthogonal to the preceding PCs and explains maximal variation in that direction. This is an unsupervised method, meaning that it does not take into account a response variable.

FDA finds an approximation of the Bayes' classifier, which is the theoretically optimal separator of the observations that pertain to different classes. In FDA we find the vector $\boldsymbol{w}$, which maximizes the following objective function

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T S_B \boldsymbol{w}}{\boldsymbol{w}^T S_W \boldsymbol{w}},$$

where $S_B = (\boldsymbol{m}_1 - \boldsymbol{m}_2)(\boldsymbol{m}_1 - \boldsymbol{m}_2)^T$ and $S_W = \sum_{i=1,2}\sum_{x \in X_i}(\boldsymbol{x} - \boldsymbol{m}_i)(\boldsymbol{x} - \boldsymbol{m}_i)^T$.

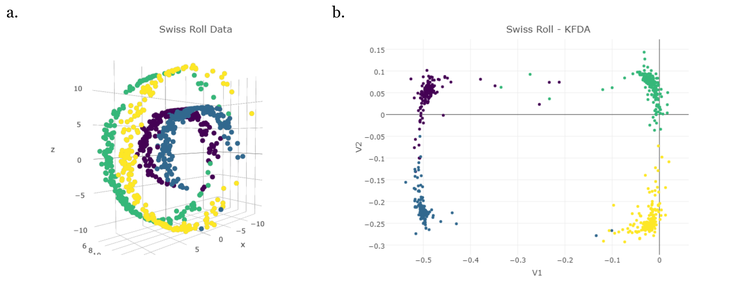
are the between class and within class scatter matrices with $m_i = \frac{1}{l_i}\sum_{j=1}^{l_i} \boldsymbol{x}_j^i$.

Simulation Studies

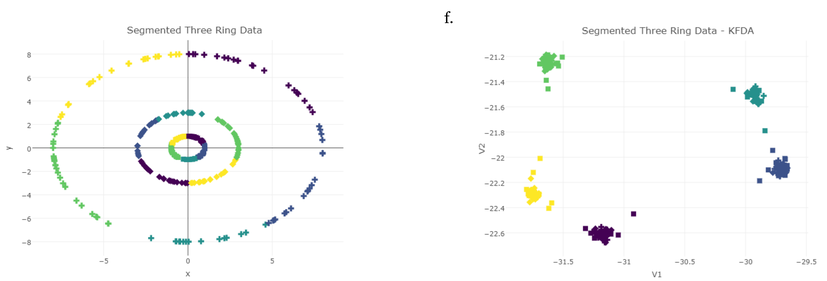
Swiss Roll Data

Segmented Concentric Ring Data

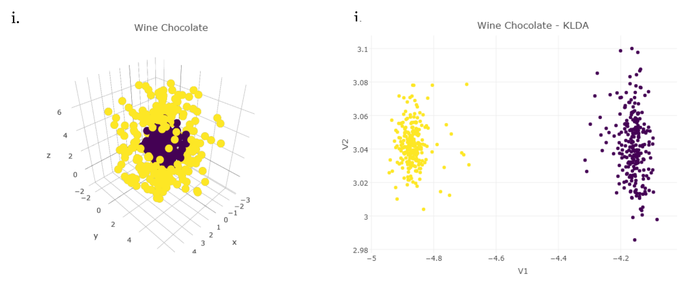
Wine Chocolate Data



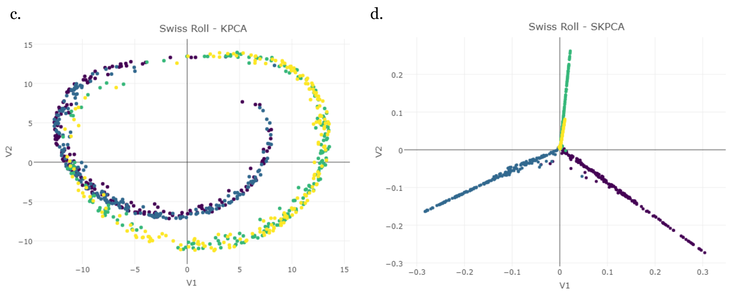
a) Simulated swiss roll data in 3-dimensions. Different colors are used to denote data instances from different classes. b) Linearly separable swiss roll data projected to 2-dimensions after application of the KFDA DR algorithm.



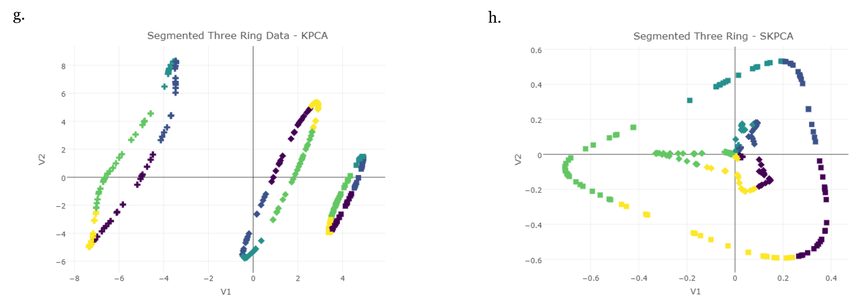
e) Simulated segmented concentric ring data in 2-dimensions. Different colors are used to denote data instances from different classes and different symbols are used to track the observations from different rings. f) Linearly separable concentric ring data projected to 2-dimensions after application of the KFDA DR algorithm.



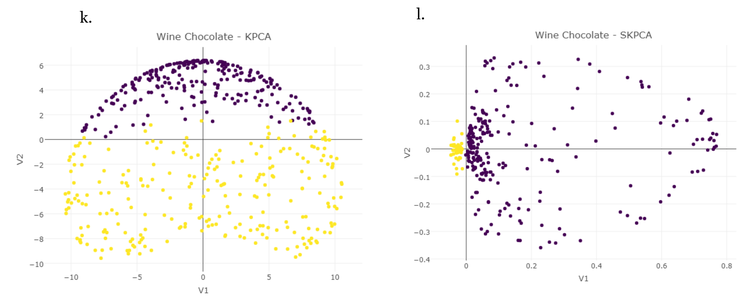
i) Simulated wine chocolate data in 3-dimensions. Different colors are used to denote data instances from different classes. j) Linearly separable wine chocolate data projected to 2-dimensions after application of the KFDA DR algorithm.



c) KPCA failed to linearly separate swiss roll data when projected to 2-dimentions. d) Linearly separable swiss roll data projected to 2-dimensions after application of the SKPCA DR algorithm.



g) KPCA failed to linearly separate concentric ring data when projected to 2-dimensions. f) SKPCA aligned the observations from common classes within each ring. The data is not linearly separable, however a better tuning of the kernel parameter γ would likely separate classes.



k) Linearly separable wine chocolate data projected to 2-dimensions after application of the KPCA DR algorithm. d) Linearly separable wine chocolate data projected to 2-dimensions after application of the SKPCA DR algorithm.

Methodology (cont.)

Non-linear Dimensionality Reduction Methods

The non-linear versions of PCA and FDA considered here depend on an implicit non-linear map $\Phi: \mathbf{R}^n \to F$ which maps elements of X to $\Phi(X)$, where F is a space of infinite dimensionality. Given a kernel function, in our case the Gaussian kernel,

$$K(\boldsymbol{x}, \boldsymbol{x}') = \Phi(\boldsymbol{x})^T\Phi(\boldsymbol{x}') = \exp(-\gamma||\mathbf{x} - \boldsymbol{x}'||^2)$$

we can efficiently compute the kernel matrix without ever explicitly finding Φ. The results given by the Gaussian kernel depend heavily on the selection of the tuning parameter γ. By replacing all occurrences of dot products in the linear dimensionality reduction methods by $\Phi(\mathbf{x})^T\Phi(\boldsymbol{x}')$, we project the data to a high-dimensional dot product space in which our data become linearly separable. In this space, we perform the linear version of the dimensionality reduction algorithm, then we are able to project the separated data back to a low-dimensional space for visualization. It is notable that the kernel-based methods are actually a generalization of the linear methods. Utilizing the polynomial kernel,

$$K(\mathbf{x}, \mathbf{x}') = \left((\boldsymbol{x}^T\boldsymbol{x}') + c\right)^d$$

with c=0 and d=1 in KPCA or KFDA is equivalent to performing linear PCA or FDA.

Acknowledgments

A cordial thank you is due to Dr. Yishi Wang, Dr. Cuixian Chen, Troy Kling, and Katherine Kempfert for their tutelage and contributions to this work. Also, thank you to the National Science Foundation (Grant #: 1659288) for establishing this undergraduate research experience and providing funding.

Conclusions

With respect to the swiss roll data, KPCA failed to produce linearly separable data, while KFDA and SKPCA produced linearly separable data. Only KFDA was able to linearly separate the segmented concentric ring data in 2-dimensions. All three DR methods were able to linearly separate the wine chocolate data. It seems that KFDA outperformed the other algorithms on all datasets. In general, KPCA and SKPCA were more sensitive to adjustments to the kernel tuning parameter. The tuning parameters, found using grid search, are not necessarily optimal. Optimal tuning parameters for KPCA and SKPCA would likely produce linearly separable projections with efficacy comparable to KFDA.

References

1. Schölkopf, B., Smola, A., & Müller, K. (1998). Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*,*10*(5), 1299-1319.
2. S. Mika, G. Ratsch, J. Weston, B. Schölkopf and K. R. Mullers, "Fisher discriminant analysis with kernels," *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No.98TH8468)*, Madison, WI, 1999, pp. 41-48.
3. Barshan, Elnaz, Ali Ghodsi, Zohreh Azimifar, and Mansoor Zolghadri Jahromi. "Supervised principal component analysis: Visualization, classification and regression on subspaces and submanifolds." Pattern Recognition44.7 (2011): 1357-371.
4. Wang, Yishi, Cuixian Chen, Valerie Watkins, and Karl Ricanek. "Modified Supervised Kernel PCA for Gender Classification." Lecture Notes in Computer Science Intelligence Science and Big Data Engineering. Image and Video Data Engineering(2015): 60-71.