Beyond the PCA: A Comprehensive Review of Dimension Reduction Techniques

Tanmay H. Kenjale, UNC Charlotte
Dr. Eliana Christou, Department of Mathematics and Statistics

UNIVERSITY OF NORTH CAROLINA CHARLOTTE

Introduction

Background

Regression analysis models the relationship between predictor variables and the response variable.

Curse of Dimensionality: as the number of predictors increases, regression analysis becomes challenging.

Dimension reduction techniques reduce the number of predictors while maintaining information.

Technique Categories

- *Supervised*: response is taken into account
- *Unsupervised*: response is *not* taken into account
- *Linear*
- *Nonlinear*

Objectives

Goals

1) Analyze several dimension reduction techniques
2) Provide a framework for comparing performances of unsupervised and supervised techniques
3) Provide recommendations for choosing a technique

Analyzed Techniques

Principal Component Analysis (PCA) [3]: unsupervised, linear

Kernel Principal Component Analysis (KPCA) [4]: unsupervised, nonlinear

Sliced Inverse Regression (SIR) [2]: supervised, linear

Sliced Average Variance Estimation (SAVE) [1]: supervised, linear

Kernel Sliced Inverse Regression (KSIR) [5]: supervised, nonlinear

Methodology

Each dimension reduction technique was tested on 4 real data sets in the following manner:

Sample Level Algorithm

1) Split data set into a 10-fold cross validation set
2) Perform each technique on the training folds
3) Estimate the dimension reduction subspace size ($\hat{d}$) for each technique:
   - For unsupervised techniques, choose the dimension that explains 60% of variation
   - For supervised techniques, perform chi-squared sequential test with $\alpha = 0.05$
4) Form the reduced predictors for each technique
5) Regress the response on the reduced predictors using a nonparametric regression model for each technique
6) Calculate the test error (RMSE) for each technique
7) Repeat Steps 1-6 for each fold and report the average $\hat{d}$ and the average RMSE for each technique

Computational time for each dimension reduction technique is also computed and averaged to compare the efficiencies of each technique.

Results

Data Set	Name	n	p
1	Boston Housing	506	13
2	Ozone	330	9

Data Set	Technique	$\hat{d}$	RMSE	Time (ms)
1	PCA	3.0	6.39	**0.88**
	KPCA	**1.5**	8.13	128.58
	SIR	3.0	5.84	5.13
	SAVE	4.0	9.04	4.10
	KSIR	3.5	**4.31**	125.25
2	PCA	2.0	4.75	**1.07**
	KPCA	1.5	5.42	53.56
	SIR	**1.0**	4.55	5.63
	SAVE	3.0	4.82	5.96
	KSIR	2.0	**3.96**	57.23

About the Data

- The first table summarizes the sample sizes (n) and number of variables (p) of 2 data sets
- The second table summarizes the results of the comparison procedure on the 2 data sets
- The smallest values in each column are bolded for each data set

Interpretations

- A lower $\hat{d}$ indicates a greater degree of dimension reduction
- A lower RMSE indicates that the dimension reduction preserves more information
- A low Time indicates that the technique executed quickly

Conclusions

	Pros	Cons
PCA	• Fastest	• Mediocre $\hat{d}$ and RMSE
KPCA	• Low $\hat{d}$	• High RMSE • Slow
SIR	• Fast • Low $\hat{d}$ • Low RMSE	
SAVE	• Fast	• High $\hat{d}$ and RMSE
KSIR	• Low $\hat{d}$ • Lowest RMSE	• Slow

Recommendations

- PCA should be tested first due to its simplicity and speed despite its lower performance
- SIR has the best combination of $\hat{d}$, RMSE, and speed
- If PCA or SIR do not perform adequately and speed is not an issue, consider KSIR

References

[1] Cook, R. D., & Weisberg, S. (1991). Sliced Inverse Regression for Dimension Reduction: Comment. Journal of the American Statistical Association, 86(414), 328–332.

[2] Li, K.-C. (1991). Sliced Inverse Regression for Dimension Reduction. Journal of the American Statistical Association, 86(414), 316–327.

[3] Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. Philosophical Magazine, 2(11):559–572.

[4] Schölkopf, B., Smola, A.J., and Müller, K.R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 10(5):1299–1319.

[5] Wu, H. (2008). Kernel Sliced Inverse Regression with Applications to Classification. Journal of computational and graphical statistics, 17:590–610.