

DOCTORAL THESIS

Dimension reduction and variable selection in regression

Wen, Songqiao

Date of Award:
2008

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

Dimension Reduction and Variable Selection in Regression

WEN Songqiao

A thesis submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Principal Supervisor: Professor ZHU Lixing

Hong Kong Baptist University

September 2008

Abstract

Consider the nonparametric regression of Y against \mathbf{X} . When the dimension of \mathbf{X} gets higher, the standard techniques such as local smoothing break down quickly because of the sparseness of the data points in any region of interest, a phenomenon often called the “curse of dimensionality”. There are essentially two approaches in literature to overcome these difficulties, the first is largely concerned with function approximation and the second is Sufficient Dimension Reduction, example of the former is additive models and example of the latter is Sliced Inverse Regression (Li (1991)).

There are essentially two parts in this thesis, which represent our efforts along the above two approaches. The first part, including chapter 2 and chapter 3, is concerned with dimension reduction with multivariate responses data, the second part, which includes chapter 4, is concerned with variable selection in additive regression models.

Consider the dimension reduction problem where both the response and the predictor are vectors. Existing estimators of this problem take one of the following routes: (1) targeting the part of the dimension reduction space that is related to the conditional mean (or moments) of the response vector, (2) pooling the estimates for the marginal dimension reduction spaces, and (3) estimating the whole dimension reduction space directly by multivariate slicing. However, the first two approaches do not fully recover the dimension reduction space, and the third is hampered by the fact that the accuracy of estimators based on multivariate slicing drops sharply as the dimension of response increases, again the “curse of dimensionality”. In chapter 2, we proposed a general method, which is called Projective Resampling method

in this thesis, to turn multivariate-response dimension reduction into univariate-response dimension reduction, and thereby avoiding the curse of dimensionality caused by slicing the multivariate response, as required in the direct extensions of the classical methods such as SIR and SAVE. The method is compared with the existing estimators by simulation, and applied to a data set. Chapter 3 represents another effort of mine on this topic. In this chapter, we proposed a new method based on the covariance matrices of predictors and the characteristic function of response variable. The merit of this method is that it can avoid the selection of turning parameters, which is often an important issue, as the performance of many methods, such as SAVE, relies heavily on the choice of slice number.

In chapter 4, we consider the problem of variable selection in additive regression models. The additive model plays an increasing important role in modern statistical analysis. Similar as in the linear regression model, how to select the significant independent variables, and obtain a simple model is also an important issue for the additive modeling. In this chapter we adapt the group variable selection technique developed by Zhou and Zhu (2007), and combine penalized regression spline technique proposed by Mammen and Van de Geer (1997) to perform variable selection and estimation for the additive models. We showed that our proposed method can select significant variables and estimate the nonparametric additive function components with optimal convergence rate simultaneously. Simulation are implemented to investigate performance of the new method.

Keywords: Central Mean Subspace, Central Subspace, Monte Carlo integration, Multivariate nonlinear regression, Sliced Average Variance Estimator, Sliced Inverse Regression; Variable selection, Penalized splines, Additive models, LASSO, Cross-Validation.

Table of Contents

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
Chapter 1 Introduction	1
1.1 Curse of Dimensionality in Regression	1
1.2 Overview of Sufficient Dimension Reduction	4
1.2.1 Sliced Inverse Regression	6
1.2.2 Sliced Average Variance Estimation	8
1.2.3 Determine the Dimension of the Central Subspace	9
Chapter 2 On a Projective Resampling Method for Dimension Reduction with Multivariate Responses	11
2.1 Introduction	11
2.2 Projective Resampling and its Application	14
2.2.1 The Method	14
2.2.2 Application	17

2.3	Properties of Projective Resampling	20
2.3.1	Fisher Consistency	20
2.3.2	\sqrt{n} -consistency and Monte Carlo effect	22
2.4	Simulation Comparisons	24
2.5	Conclusion	32
2.6	Proof of Theorems	33
Chapter 3 Covariance Methods on Dimension Reduction in Regres-		
sions with Multivariate Responses		38
3.1	Introduction	38
3.2	Methodology Development	40
3.3	Estimation and Asymptotic Normality	43
3.4	Application	44
3.5	Simulation	48
3.6	Discussion	54
3.7	Proof of Theorems	55
Chapter 4 Variable Selection in Additive Regression Models		57
4.1	Introduction	57
4.2	Methodology	61
4.2.1	Group Variable Selection Methods	61
4.2.2	Penalized Regression Splines	62
4.2.3	Group Variable Selection for Additive Models	63
4.2.4	Asymptotic Theory	64
4.2.5	Algorithms and Turning Parameters Selection	66
4.3	Numerical Studies	68
4.4	Discussion	74
4.5	Proof of Theorems	75
Bibliography		79

