



Doctoral Thesis

High-dimensional regression problems with special structure

Author(s):

Meier, Lukas D.

Publication Date:

2008

Permanent Link:

<https://doi.org/10.3929/ethz-a-005784712> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 18129

High-Dimensional Regression Problems with Special Structure

A dissertation submitted to
ETH ZURICH

for the degree of
Doctor of Sciences

presented by
LUKAS DIETER MEIER
Dipl. Math. ETH
born August 9, 1979
citizen of Unterkulm AG

accepted on the recommendation of
Prof. Dr. Peter Bühlmann, examiner
Prof. Dr. Sara van de Geer, co-examiner

2008

Abstract

High-dimensional variable selection has received tremendous attention in the last decade. Sparse estimators like the Lasso (Tibshirani, 1996) have proven to be powerful for high-dimensional settings, both from a theoretical and a practical point of view. In this thesis we propose several extensions of the Lasso (and of variants thereof) when either additional structure is available in the data or when more flexibility with respect to sparsity or functional modeling is of interest.

First, we consider the Group Lasso (Yuan and Lin, 2006). The Group Lasso is an extension of the Lasso to do variable selection on (predefined) groups of variables in linear regression models. The estimates have the attractive property of being invariant under groupwise orthogonal reparametrizations. We extend the Group Lasso to logistic regression models and present an efficient algorithm, especially suitable for high-dimensional problems, which can also be applied to generalized linear models to solve the corresponding convex optimization problem. The Group Lasso estimator for logistic regression is shown to be statistically consistent even if the number of predictors is much larger than sample size but with sparse true underlying structure. We further use a two-stage procedure which aims for sparser models than the Group Lasso, leading to improved prediction performance for some cases. Moreover, due to the two-stage nature, the estimates can be constructed to be hierarchical. The methods are used on simulated and real datasets about splice site detection in DNA sequences.

When a series of (related) linear models has to be estimated it is often appropriate to combine the different data-sets to construct more efficient estimators. We use ℓ_1 -penalized estimators like the Lasso or

the Adaptive Lasso (Zou, 2006) which can simultaneously do parameter estimation and model selection. We show that for a time-course of high-dimensional linear models the convergence rates of the Lasso and of the Adaptive Lasso can be improved by combining the different time-points in a suitable way. Moreover, the Adaptive Lasso still enjoys oracle properties and consistent variable selection. The finite sample properties of the proposed methods are illustrated on simulated data and on a real problem of motif finding in DNA sequences.

Often, the assumption of a linear relationship between the response and the predictors is too restrictive. We propose a new sparsity-smoothness penalty for high-dimensional generalized additive models. The combination of sparsity and smoothness is crucial for mathematical theory as well as performance for finite-sample data. We present a computationally efficient algorithm, with provable numerical convergence properties, for optimizing the penalized likelihood. Furthermore, we provide oracle results which yield asymptotic optimality of our estimator for high-dimensional but sparse additive models. Finally, an adaptive version of our sparsity-smoothness penalized approach yields large additional performance gains.

Assigning significance in high-dimensional regression is challenging. Most computationally efficient selection algorithms cannot guard against inclusion of noise variables. Asymptotically valid p-values are not available. An exception is a recent proposal by Wasserman and Roeder (2008) which splits the data into two parts. The number of variables is then reduced to a manageable size using the first split, while classical variable selection techniques can be applied to the remaining variables, using the data from the second split. This yields asymptotic error control under minimal conditions. It involves, however, a one-time random split of the data. Results are sensitive to this arbitrary choice: it amounts to a “p-value lottery” and makes it difficult to reproduce results. Here, we show that inference across multiple random splits can be aggregated, while keeping asymptotic control over the inclusion of noise variables. In addition, the proposed aggregation is shown to improve power, while reducing the number of falsely selected variables substantially.

The Adaptive Lasso (Zou, 2006) is a two-stage procedure. Based on an initial estimator, a weighted penalty function is constructed. The main idea is to penalize the more important variables less to remove

bias. We discuss the idea of successively applying such a weighted penalization scheme in order to get sparser solutions without losing a lot of prediction performance.

Zusammenfassung

Das Problem der Variablenwahl in hochdimensionalen Räumen ist in den letzten Jahren ausgiebig erforscht worden. Dünn besetzte Schätzer wie Lasso (Tibshirani, 1996) und Erweiterungen davon haben sich sowohl in der Praxis wie auch in der Theorie für hochdimensionale Probleme als sehr geeignet erwiesen. Diese Dissertation untersucht diverse Erweiterungen des Lasso-Schätzers für Situationen, wo zusätzliche Struktur in den Daten enthalten ist oder für den Fall, dass mehr Flexibilität bezüglich der Modellierung und der Anzahl gewählter Variablen erwünscht ist.

Als erstes betrachten wir den Group Lasso Schätzer (Yuan and Lin, 2006). Dieser ist eine Erweiterung des gewöhnlichen Lasso Schätzers und ermöglicht in linearen Regressionsproblemen die Variablenwahl bezüglich (vorher definierten) Gruppen. Der Schätzer besitzt die attraktive Eigenschaft, dass er invariant unter gruppenweise orthogonalen Umparametrisierungen ist. Wir verwenden den Group Lasso Schätzer für logistische Regressionsprobleme und stellen einen effizienten Algorithmus vor, der speziell für hochdimensionale Probleme geeignet ist. Zudem kann man die Idee auch für verallgemeinerte lineare Regressionsprobleme verwenden. Wir zeigen, dass der Group Lasso Schätzer konsistent ist, selbst wenn die Anzahl Prädiktoren viel grösser als die Stichprobengrösse ist, aber das zugrundeliegende Modell dünn besetzt ist. Ferner verwenden wir ein Zwei-Stufen Verfahren, das dünn besetztere Lösungen liefert und oft auch bessere Vorhersageeigenschaften hat. Zudem ermöglicht es die Anpassung hierarchischer Modelle. Wir verwenden die Methoden sowohl bei simulierten Daten wie auch für das Problem der Vorhersage von Splice-Sites auf der menschlichen DNA.

Als weiteres Problem betrachten wir eine Folge von (ähnlichen) linearen Modellen. Oft ist es angebracht, die verschiedenen Datensätze zu kombinieren, um einen effizienteren Schätzer zu erhalten. Wir betrachten auch hier ℓ_1 -bestrafte Schätzer wie Lasso oder Adaptive Lasso (Zou, 2006), die sowohl Parameterschätzung wie auch Modellwahl durchführen. Wir zeigen, dass für eine Serie von hochdimensionalen linearen Regressionsproblemen die Konvergenzrate für Lasso und Adaptive Lasso verbessert werden kann, wenn man die verschiedenen Zeitpunkte passend miteinander kombiniert. Adaptive Lasso hat zusätzlich immer noch “Orakel-Eigenschaften” und ist bezüglich der Modellwahl konsistent. Die Eigenschaften der Schätzer werden anhand von simulierten Daten sowie anhand der Suche von DNA-Bindestellen von Transkriptionsfaktoren illustriert.

Die Annahme eines linearen Zusammenhangs zwischen der Zielvariable und den erklärenden Variablen kann in gewissen Fällen zu einschränkend sein. Wir schlagen deshalb einen neuen Bestrafungsterm vor, der es erlaubt, mit hochdimensionalen (verallgemeinerten) additiven Modellen zu arbeiten. Die Idee besteht darin, den Bestrafungsterm so zu wählen, dass sowohl für dünn besetzte Lösungen gesorgt wird, aber gleichzeitig auch die Glattheit der Funktionen geregelt wird. Wir präsentieren einen effizienten Algorithmus, der die bestrafte Likelihood minimieren kann. Zudem zeigen wir, dass der Schätzer “Orakel-Eigenschaften” hat im Falle von hochdimensionalen aber dünn besetzten additiven Modellen. Schlussendlich zeigen wir, dass eine adaptive Version des Schätzers nochmals grosse Verbesserungen bringen kann.

P-Werte sind oft nicht verfügbar bei Verfahren für die Variablenwahl in hochdimensionalen Regressionsproblemen. In Wasserman and Roeder (2008) wird vorgeschlagen, dass man die Daten in zwei disjunkte Mengen aufteilen soll. Zuerst wird mit der ersten Teilmenge die Anzahl Variablen reduziert. Basierend auf der zweiten Teilmenge kann dann mit klassischen Verfahren die Variablenwahl verfeinert werden. Im Speziellen kann man für die im ersten Schritt gewählten Variablen p-Werte angeben und asymptotisch die Wahrscheinlichkeit kontrollieren, dass mindestens eine Variable fälschlicherweise ausgewählt wird, obwohl sie gar nicht im Modell ist. Um die Abhängigkeit von einer einzelnen Datenaufteilung zu reduzieren, führen wir obiges Verfahren mehrmals durch. Wir zeigen, dass man die daraus resultierenden p-Werte mittels Quantilen aggregieren kann, so dass die asymptotische Kontrolle über

die Fehlerrate gewährleistet bleibt.

Der Adaptive Lasso Schätzer (Zou, 2006) ist ein Zwei-Stufen Verfahren. Basierend auf einem Initialschätzer wird eine gewichtete Bestrafungsfunktion konstruiert. Wir diskutieren die Idee, diese Methode zu einem Mehrstufen-Verfahren zu erweitern, um noch dünn besetztere Lösungen zu erhalten, ohne gross an Vorhersage-Eigenschaften zu verlieren.