Diss. ETH No. 14696

Nonparametric regression penalizing deviations from additivity

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Mathematics

presented by
MICHAEL MARKUS STUDER
Dipl. Math. ETH
born September 1, 1972
citizen of Maschwanden ZH

accepted on the recommendation of
Prof. Dr. H. R. Künsch, examiner
Prof. Dr. P. Bühlmann, co–examiner
Prof. Dr. T. Gasser, co–examiner
PD. Dr. B. Seifert, co–examiner

2002

Abstract

Non–parametric regression in higher dimensions suffers from the 'curse of dimensionality'. The assumption of an arbitrary multivariate smooth regression function is not sufficiently specific for practice, as there exists no acceptable compromise between variability and oversmoothing. This is the reason for using simpler models, even if they do not hold. Additive models have more restrictive assumptions and are hence easier to estimate. As a drawback, there is an additional model–bias as compared to the full model. We are investigating a *continuous compromise* between these two models, containing the above mentioned extremes as special cases.

This is achieved via a local linear estimator with penalty on the deviation from the additive model. We evaluate the estimator on a grid of output points. The subset of all possible parameter values, corresponding to the additive model, is denoted as additive subspace. The simultaneous local linear estimator for all output points is the solution of a quadratic minimization problem with many variables but simple structure. A multiple (parameter R) of the squared Euclidean distance to the additive subspace is added as a penalty term to the minimization problem. The ordinary local linear estimator corresponds to the case $R = 0$. Formally, $R = \infty$ corresponds to a local linear estimator restricted to the additive model, and the resulting estimator is asymptotically optimal if the grid is replaced by a continuum (Mammen, Linton, and Nielsen, 1999). The estimator with penalty on the non–additive component is continuous in R — for $R = 0$ this applies only to the output points where the local linear estimator is unique. Accordingly, we have a parametric class of estimators, whose extremes are competitive for the corresponding classes of regression functions.

This indicates that the new method has the potential to improve the original estimators. The estimator was applied to simulated data for illustration. A simple parameter selection criterion is the improved Akaike information criterion (AIC) proposed by Hurvich, Simonoff, and Tsai (1998). The optimal parameters were in general neither 0 nor ∞. Extremes were detected by AIC, hence nothing was lost in these cases.

The special structure of the minimization problem allows reducing matrix size. This is necessary for computation. Furthermore, the conversion improves our understanding.

An attractive property of the estimator with penalty of the non–additive components is that the estimation is a data adaptive continuous compromise between multivariate local linear and additive estimation.

On the one hand, the local linear estimator has good asymptotic properties and is easily transcribed to higher dimensions. On the other hand, regularization is required in practice. A byproduct of penalizing is that the estimator is regularized even for small penalties.

Keywords: non–parametric estimation, additive models, shrinkage, model choice, curse of dimensionality, regularization, optimal bandwidth.

Zusammenfassung

Nichtparametrische Regressionen in höheren Dimensionen sind vom 'Fluch der Dimension' betroffen. Die Annahme einer beliebigen multivariaten, glatten Regressionsfunktion ist zu unspezifisch für die Praxis, weil kein befriedigender Kompromiss zwischen Variabilität und Überglätten existiert. Deshalb werden einfachere Modelle verwendet, selbst wenn diese nicht gelten. Additive Modelle verwenden restriktivere Annahmen und sind daher besser schätzbar. Ein Nachteil gegenüber dem vollen Modell ist der zusätzliche Modell–Bias. Wir untersuchen einen *stetigen Kompromiss* zwischen den beiden Modellen, der die beiden oben erwähnten Extreme als Spezialfälle enthält.

Dies geschieht mittels einem lokal linearen Schätzer mit Bestrafung der Abweichung vom additiven Modell. Wir betrachten den Schätzer auf einem Gitter von Ausgabepunkten. Die Teilmenge aller möglichen Parameterwerte, die dem additiven Modell entsprechen, wird als additiver Unterraum bezeichnet. Der gemeinsame lokal lineare Schätzer für alle Ausgabepunkte ist die Lösung eines quadratischen Minimierungsproblems mit vielen Variablen aber einer einfachen Struktur. Ein Vielfaches (Parameter R) des quadrierten euklidischen Abstands zum additiven Unterraum wird zum Minimierungsproblem als Bestrafungsterm hinzugefügt. Der unmodifizierte lokal lineare Schätzer entspricht dem Fall $R = 0$. Dem Fall $R = \infty$ entspricht ein lokal linearer Schätzer mit Restriktion auf das additive Modell, und der resultierende Schätzer ist asymptotisch optimal, falls das Gitter durch ein Kontinuum ersetzt wird, wie Mammen, Linton und Nielsen (1999) gezeigt haben. Der Schätzer mit Bestrafung der nichtadditiven Komponente ist stetig in R — für $R = 0$ jedoch nur an den Ausgabepunkten, an denen der lokal lineare Schätzer eindeutig ist. Wir haben also eine parametrische Klasse von Schätzern, wobei die beiden Extremfälle brauchbare Schätzer sind.

Dies ist ein Hinweis darauf, dass diese Methode Potential zur Verbesserung der ursprünglichen Schätzer hat. Zur Illustration wurde der Schätzer auf simulierte Daten angewandt. Als einfaches Kriterium zur Parameterwahl wurde das verbesserte Akaike Informations Kriterium (AIC) von Hurvich, Simonoff und Tsai (1998) verwendet. Die optimalen Parameter waren im allgemeinen weder 0 noch ∞. Die Extremfälle wurden von AIC erkannt, so dass auch hier nichts verloren geht.

Die spezielle Struktur der Minimierungsproblems erlaubt eine Reduktion der Grösse der Matrizen, was angesichts der grossen Anzahl Parameter für die rechentechnische Umsetzung notwendig ist. Diese Umformung führt zu einen besseren Verständnis der theoretischen Eigenschaften.

Eine attraktive Eigenschaft des Schätzers mit Bestrafung der nichtadditiven Komponente ist, dass die Schätzung ein datenadaptiver stetiger Kompromiss zwischen der multivariaten lokal linearen Schätzung und einer additiven Schätzung ist. Die Gewichtung hängt davon ab, wieviele Daten lokal vorhanden sind.

Der lokal lineare Schätzer hat einerseits gute asymptotische Eigenschaften und ist einfach auf höhere Dimensionen übertragbar. Andererseits ist in der Praxis eine Regularisierung notwendig. Ein Nebenprodukt ist die Regularisierung des Schätzers schon für kleine Bestrafungsterme.