

15. Mai 2007

Einladung

Im Kolloquium Stochastik für die Praxis spricht

Prof. Dr. Torsten Hothorn
Ludwig – Maximilians - Universität München

über das Thema

Schätzung von Regressionsmodellen mit Boosting-Verfahren: Theorie und Praxis

Der Vortrag findet statt am

Freitag, 01. Juni 2007 um 14:15 Uhr

im

Seminarraum der Stochastik, Maschmühlenweg 8 - 10

Es laden ein: die Dozenten des Zentrums für Statistik.

Abstract:

Die Schätzung von Regressionsmodellen dient im Wesentlichen zwei Zielen: der Interpretation des Zusammenhangs zwischen Einflussgrößen und Zielgröße sowie der Vorhersage der Zielgröße für interessante Konfigurationen der Einflussgrößen. Auch wenn für den Zweck der Vorhersage komplexe, nicht mehr interpretierbare Modelle nützlich sein können (sogenannte black boxes), ist für die Interpretation des Regressionszusammenhanges ein möglichst einfaches und sparsames Modell notwendig.

Um ein solches Modell schätzen zu können, bedarf es Verfahren zur Selektion von relevanten Einflussgrößen sowie zur Modellierung von potentiell vorhandenen nicht-linearen Zusammenhängen - unter Umständen auch in Situationen, in denen die Anzahl der Einflussgrößen die Anzahl der Beobachtungen um ein Vielfaches übersteigt, wie z.B. in der Analyse von Gen-Expressionsdaten, und klassische Verfahren, wie etwa lineare oder additive Modelle, nicht mehr anwendbar sind.

Mit dem Gradienten-Boosting, einem Verfahren zur Schätzung von Regressionsmodellen mittels empirischer Risikominimierung, können die oben skizzierten Probleme mithilfe eines übergreifenden theoretischen Konzeptes behandelt werden. Es wird eine neue Klasse von flexiblen Regressionsmodellen für Überlebenszeiten vorgestellt, deren Schätzung mit einem eigens dafür entwickelten Boosting-Algorithmus erfolgt (Hothorn et al., 2006). Für die Auswahl eines geeigneten Modells aus einer Menge von konkurrierenden Modellen werden Benchmark-Experimente vorgestellt und angewandt (Hothorn et al., 2005).

Die praktische Bedeutung der beschriebenen Methoden wird anhand von Regressionsanalysen zu verschiedenen Fragestellungen aus der Ernährungswissenschaft, Ökologie und Medizin diskutiert. Um den Transfer von neuen Techniken in die statistische Praxis zu ermöglichen, sind frei verfügbare Implementierungen notwendig. Die Analyse der Beispielanwendungen erfolgt entsprechend mit dem Zusatzpaket `mboost` (Hothorn and Bühlmann, 2006) zum R-System für statistisches Rechnen.

Literatur

Torsten Hothorn and Peter Bühlmann. Model-based boosting in high dimensions. *Bioinformatics*, 22(22):2828–2829, 2006. doi: 10.1093/bioinformatics/btl462.

Torsten Hothorn, Friedrich Leisch, Achim Zeileis, and Kurt Hornik. The design and analysis of benchmark experiments. *Journal of Computational and Graphical Statistics*, 14(3):675–699, 2005. doi: 10.1198/106186005X59630.

Torsten Hothorn, Peter Bühlmann, Sandrine Dudoit, Annette Molinaro, and Mark J. van der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, Jul 2006. doi: 10.1093/biostatistics/kxj011.