

LMU München, Institut für Statistik, Seminar für angewandte Stochastik

Seminar im Wintersemester 2014/2015

Regularisierungstechniken in Regression und Latent Trait Modellen

Prof. Dr. G. Tutz;

M. Berger, M.Sc.; Dipl.-Stat. M. Oelker; Dipl.-Stat. W. Pössnecker;
G. Schauburger, M.Sc.; Dipl.-Stat. M. Schneider;

In Regressionsproblemen mit einer großen Anzahl an Einflussgrößen genauso wie in Latent Trait Modellen ist es notwendig die relevanten Variablen bzw Wirkungsstrukturen zu identifizieren, da die übliche Maximum Likelihood-Schätzung oft nicht durchführbar oder zumindest nicht optimal ist. Es werden unterschiedliche Ansätze betrachtet, wie sich die Prädiktorenstruktur spezifizieren lässt und wie Regularisierungstechniken einsetzbar sind, um Parameter schätzbar zu machen und relevante Strukturen zu identifizieren. Ein wesentlicher Schwerpunkt ist die Variablenselektion. Viele der betrachteten Verfahren zielen auf eine Kombination von Variablenselektion und Verbesserung der Modellperformance ab.

Das Seminar richtet sich an Studierende im Bachelor- und Masterstudiengang Statistik. Als Hintergrund-Literatur, in dem viele der Verfahren kurz skizziert sind, dient das im Netz verfügbare Buch:

Hastie, T., Tibshirani, R. & Friedman, J. (2009): The Elements of Statistical Learning – Data Mining, Inference and Prediction, 2nd Edition, New York: Springer.

Weitere Literatur zu den einzelnen Themen wird im Seminar bekannt gegeben.

Das Seminar findet dienstags von 16 – 18 Uhr im Seminarraum des Instituts für Statistik, Ludwigstr. 33, statt. Die Anmeldung zum Seminar erfolgt über das LSF. Erster Seminar-Termin ist Dienstag, der 7.10.2014 (Einführung, Themenvergabe etc.). Weitere Informationen finden Sie auf der Seminar-Homepage:

Themenvorschläge und Literatur:

- Ridge Regression
 - Hoerl & Kennard (1970): Ridge regression: biased estimation for nonorthogonal problems, *Technometrics* 12.
 - Nyquist (1991): Restricted estimation of generalized linear models, *Journal of Applied Statistics* 40.
- Lasso
 - Tibshirani (1996): Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society B* 58.
- Lasso in GLMs
 - Park & Hastie (2007): L1-regularization path algorithm for generalized linear models, *Journal of the Royal Statistical Society B* 69.
- Grouped Lasso
 - Yuan & Lin (2006): Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society B* 68.
 - Meier et al. (2008): The group lasso for logistic regression, *Journal of the Royal Statistical Society B* 70.
- Fused Lasso
 - Tibshirani et al. (2005): Sparsity and smoothness via the fused lasso, *Journal of the Royal Statistical Society B* 67.
- Elastic Net
 - Zou & Hastie (2005): Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society B* 67.
- Least Angle Regression
 - Efron et al. (2004): Least angle regression, *The Annals of Statistics* 32.
- SCAD
 - Fan & Li (2001): Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association* 96.

- Variable Selection in Categorical Regression
 - Tutz, G., Pössnecker, W. and Uhlmann, L. (2012): Variable Selection in General Multinomial Logit Models. *Technical Report 126, Department of Statistics LMU.*
- Rasch Modell - Testen und Schätzen
 - Fischer & Molenaar (1995): Rasch Models. Foundations, Recent Developments, and Applications. *Springer, New York*
 - Strobl (2010): Das Rasch Modell: Eine verständliche Einführung für Studium und Praxis, *Rainer Hampp Verlag*
- Rekursive Partitionierung für Rasch Modelle
 - Strobl et al. (2013): Rasch trees: A new method for detecting differential item functioning in the Rasch model, *Psychometrika*, 1-28
- Lasso in Rasch Models
 - Tutz & Schaubberger (2014): A Penalty Approach to Differential Item Functioning in Rasch Models. *Psychometrika*, to appear.
- Lasso in Paired Comparison Models
 - Tutz & Schaubberger (2014): Extended Ordered Paired Comparison Models with Application to Football Data from German Bundesliga, *Technical Report 151, Department of Statistics LMU*
- Trees and Random Forests
 - Breiman et al. (1984): *Classification and regression trees*, Wadsworth and Brooks, Monterey CA.
 - Breiman (2001): Random forests, *Machine Learning* 45.
 - Tutz (2012): *Regression for Categorical Data*, Cambridge University Press.
- Algorithmus-basierte Variablenselektion in Regression und Klassifikation
 - Friedman (2008): Fast sparse regression and classification, *Proceedings of the 23rd International Workshop on Statistical Modelling.*

- L_2 -Boosting
 - Bühlmann & Yu (2003): Boosting with the L_2 loss: regression and classification, *Journal of the American Statistical Association* 98 (462), 324-339.
 - Bühlmann & Hothorn (2007): Boosting Algorithms: Regularization, Prediction and Model Fitting, *Statistical Science*, 477-505.
- Likelihood-based Boosting
 - Tutz & Binder (2006): Generalized additive modelling with implicit variable selection by likelihood based boosting. *Biometrics*, 62, 961-971.