
Remagen, den 25.11.2017

Mastermodul im Sommersemester 2018

Ausgewählte Themen aus der Statistik:

Regressionsmodelle – klassisch und hochdimensional

Lernziele:

Die Studierenden können die Annahmen von Regressionsmodellen umfassend prüfen. Sie können spezielle Fragestellungen in Regressionsmodellen bearbeiten, auch im Fall von heteroskedastischen und korrelierten Residuen. Sie können die wesentlichen Prädiktorvariablen im Modell identifizieren. Für höherdimensionale und hochdimensionale Modelle können Sie Verfahren wie Ridge-Regression, Lasso und Boosting anwenden und die Ergebnisse sachgerecht interpretieren.

Inhalte:

Schätzen der Parameter im linearen Regressionsmodell, Residuenanalyse, multivariate Tests und Konfidenzbereiche, Prognoseintervalle.

Das Modell mit heteroskedastischen und korrelierten Fehlertermen, Regression mit autokorrelierten Daten, Durbin-Watson-Test.

Bayesianisches Regressionsmodell, Gibbs Sampler zur Bestimmung der posterioren Verteilung der Parameter.

Variablenselektionsmaße für die Modellgüte (AIC etc.), Kollinearitäts- und Ausreißerdiagnostik.

Ridge Regression als einfachstes Penalisiertes-Verfahren. Wahl des Penalisierungsparameters. Lasso (least absolute shrinkage and selection operator), eine Methode die ähnlich wie Ridge Regression aussieht, aber gleichzeitig eine Variablenselektion durchführt und auf hochdimensionale Daten anwendbar ist. LARS (least angle regression) als sehr effizienter Lasso-Schätzer.

Boosting, ein Verfahren aus dem Machine Learning, das mit sehr vielen kleinen Schritten ein flexibles Modell anpassen kann und dabei eine Überanpassung vermeidet.

Vorkenntnisse: Statistik 2.

Literatur:

Fahrmeir L, Kneib T, Lang S: Regression, Springer, Berlin 2007

Neter J, Kutner MH, Nachtsheim CJ, Wasserman W: Applied Linear Statistical Models, McGraw Hill, Boston, 1996.

Hastie T, Tibshirani R, Freedman J: The Elements of Statistical Learning, Springer Berlin, 2008. Papers werden für die zweite Semesterhälfte zur Verfügung gestellt.
