

Hauptseminar Erweiterungen des linearen Regressionsmodells und genomische Anwendungen in der Biomedizin, WS 2014/15

Liste der Vorträge

Harald Binder und Matthias Birkner*

Stand: 7.8.2014

Termin: Mo. 8:30-10h, Beginn in der 2. Vorlesungswoche

1. Zum linearen Modell, klassischer kleinste-Quadrate- und Maximum-Likelihood-Schätzung

(Matthias Birkner)

2. Generalisiertes lineares Modell mit hochdimensionaler Anwendung in R

(Harald Binder)

3. Berechnung von regularisierten Schätzern

Vorstellung von Berechnungsverfahren zur Bestimmung von Lösungen für lasso, ridge regression und elastic net via koordinatenweise (konvexe) Optimierung für den linearen Fall und für den Fall logistischer verallgemeinerter Regression; Diskussion der Korrektheit; Illustration anhand von Beispiel(en) mittels R-Paket [glmnet]

Quellen: [FHT10], ggfs. Auszüge aus Kap. 2 von [BvdG11], zur Einführung des verallgemeinerten linearen Modells auch Kap. 2.1 und 2.2 von [FT01], für Beispieldaten ggfs. auch [R02]

(Lukas Lippa)

4. Konsistenzeigenschaften, adaptives lasso

Diskussion möglicher Verzerrung des lasso-Schätzers, Vorstellung des adaptive lasso zur Behebung; Konsistenzeigenschaften des (adaptive) lasso via sog. oracle inequalities und Diskussion der Asymptotik, Optimalitätseigenschaften (mit Beweis(skizzen)); Illustration anhand von simulierten Datensätzen (mit R und [glmnet])

Quellen: [Z06] und ggfs. Referenzen dort, Auszüge aus Kap. 2 sowie aus Kap. 6 und 7 von [BvdG11]

(Kristina Kaucher)

5. Das elastic net und Gruppierung korrelierter Prädiktoren

Diskussion des Verhaltens des lasso bei stark korrelierten Prädiktoren; Vorstellung des elastic net, das in diesem Fall die Schätzwerte der Koeffizienten gleichmäßig(er) auf die Gruppen verteilt; Diskussion der Eigenschaften des elastic net, ggfs. mit Beweis(en); Illustration anhand von simulierten und ggfs. auch realen Datensätzen (mit R und [glmnet])

Quellen: [ZH05], Auszüge aus Kap. 2 von [BvdG11], für Beispieldaten ggfs. auch [R02]

(Bettina Wiebe)

*birkner@mathematik.uni-mainz.de

6. Signifikanztests im hochdimensionalen Kontext I : Überblick

Diskussion von Ansätzen für statistische Tests im hochdimensionalen Kontext, ob gewisse Koeffizienten im wahren Modell verschwinden: Stichworte stability selection, subsample splitting, Eigenschaften; Illustration anhand von simulierten Datensätzen (mit R und [glmnet]); ggfs. auch Einführung in graphische Modelle

Quellen: [BRK11] und ggfs. Referenzen dort, Auszüge aus Kap. 2 sowie ggfs. aus Kap. 7, 10 von [BvdG11], (Thomas Schäfer)

7. Signifikanztests im hochdimensionalen Kontext II : Ein neuer Ansatz

Ein Ansatz, das statistische Testproblem für eine ausgewählte Einflussgröße zu lösen, indem der verzerrende Effekt der anderen Einflussgrößen modelliert wird

Quelle: [B13] und ggfs. Referenzen dort
(Stella Preußler)

8. Regularisierte Schätzung für Ereigniszeiten

Knappe Einführung in Überlebenszeitmodellierung; hochdimensionale regularisierte Schätzung der Koeffizienten, die die hazard rate bestimmen, anhand von zensierten Beobachtungen, à la lasso; Konsistenzeigenschaften; ggfs. Illustration anhand simulierter Daten und/oder der Daten aus [R02]

Quelle: [LL13] und Referenzen dort (ggfs. Auszüge aus Kap. 2 sowie aus Kap. 6 und 7 von [BvdG11] für begleitenden Hintergrund), Basisliteratur für Überlebenszeitmodellierung z.B. [ABG08]

(Paul Schulz)

9. Sure Independent Screening für Ereigniszeiten

Reduktion der Kandidaten für Einflussgrößen in ultrahochdimensionalen Problemen unter Annahme von (approximativer) Unkorreliertheit

Quelle: [GS13] und Referenzen dort (begleitende Hintergrundinformation wie im Vortrag Regularisierte Schätzung für Ereigniszeiten)

(Jan Weimer)

10. Lasso mit Interaktionstermen

Vorstellung des hierarchischen lasso für hochdimensionale Regression mit Paarinteraktion, Berechnungsverfahren und Eigenschaften (insbesondere: Erhalt der „hierarchischen“ Struktur, d.h. ein Paarterm wird nur dann in das Modell aufgenommen, wenn zugleich jeder der Terme einzeln vorkommt), Illustration anhand von simulierten Datensätzen (mit R und [hierNet])

Quelle: [BTT13]

(Sebastian Franz)

Anmerkungen

Die Monographien [BvdG11], [HTF11], [FT01] sind generelle Quellen, die insgesamt genommen weit über den Inhalt des Seminars hinausgehen. Sie werden in der Handbibliothek der Arbeitsgruppe Stochastik verfügbar sein.

Literatur

[ABG08] Odd Aalen, Ornulf Borgan, Hakon Gjessing, Survival and Event History Analysis: A Process Point of View, Springer, (2008).

[BTT13] Jacob Bien, Jonathan Taylor, Robert Tibshirani, A lasso for hierarchical interactions, Annals of Statistics 41(3), 1111–1141, (2013).

- [B13] Peter Bühlmann, Statistical significance in high-dimensional linear models, *Bernoulli*, 19(4), 1212–1242, (2013).
- [BRK11] Peter Bühlmann, Philipp Rütimann, Markus Kalisch, Controlling false positive selections in high-dimensional regression and causal inference. *Statistical Methods in Medical Research*, 22(5), 466–492, (2011).
- [BvdG11] Peter Bühlmann, Sara van de Geer, *Statistics for High-Dimensional Data*, Springer, (2011).
- [FT01] Ludwig Fahrmeir, Gerhard Tutz, *Multivariate statistical modelling based on generalized linear models*. 2nd Ed, Springer, (2001).
- [FHT10] Jerome Friedman, Trevor Hastie, Robert Tibshirani, Regularization paths for generalized linear models via coordinate descent, *Journal of Statistical Software* 33(1), (2010).
- [glmnet] Jerome Friedman, Trevor Hastie, Noah Simon, Rob Tibshirani, glmnet: Lasso and elastic-net regularized generalized linear models, R package, <http://cran.r-mirror.de/web/packages/glmnet/index.html>
- [GS13] Anders Gorst-Rasmussen, Thomas Scheike, Independent screening for single-index hazard rate models with ultrahigh dimensional features, *Journal of the Royal Statistical Society: Series B*, 75(2), 217-245, (2013).
- [HTF11] Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition, Springer, (2009).
Auch elektronisch erhältlich unter <http://statweb.stanford.edu/~tibs/ElemStatLearn/>
- [hierNet] Jacob Bien, Robert Tibshirani, hierNet: A Lasso for Hierarchical Interactions, R package, <http://cran.r-mirror.de/web/packages/hierNet/index.html>
- [LL13] Wei Lin, Linchi Lv, High-Dimensional sparse additive hazards regression, *Journal of the American Statistical Association*, 108(501), 247-264, (2013).
- [R02] Andreas Rosenwald et al, The use of molecular profiling to predict survival after chemotherapy for diffuse large-b-cell lymphoma, *New England Journal of Medicine* 346(25), 1936–1947, (2002).
Die Daten sind elektronisch erhältlich unter <http://llmpp.nih.gov/DLBCL/>
- [ZH05] Hui Zou, Trevor Hastie, Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society B*, 67(2), 301–320, (2005).
- [Z06] Hui Zou, The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, 101(476), 1418–1429, (2006).