

Quantitative Tools – Advanced Econometrics

Kursankündigung

M.Sc. Betriebswirtschaftslehre

Prof. Dr. A. Tarassow
Technische Hochschule Brandenburg

Stand: Oktober 2024

Kursinformationen

- **Vorlesung:** 2 SWS
- **Übung:** 2 SWS
- **Abschlussprüfung:** Klausur oder Projektarbeiten
- **Kreditpunkte:** 6 ECTS für 4 SWS

Kursleiter

Prof. Dr. A. Tarassow

Magdeburger Straße 50, A.3.0.4 (WWZ)

tarassow@th-brandenburg.de

Sprechstunde: auf Anfrage.

Kursüberblick

Der Kurs vermittelt fortgeschrittene statistische Methoden als auch maschinelle Lernverfahren mit praktischer Anwendung im Bereich der Wirtschaftswissenschaften. Der Schwerpunkt liegt auf der Implementierung moderner Verfahren aus dem Bereich des überwachten als auch unüberwachten Lernens im wirtschaftswissenschaftlichen Kontext. Besonderes Augenmerk wird auf das Domänenverständnis (*data literacy*), die praktische Umsetzung, Datenvisualisierung und Ergebnispräsentation gelegt. In den Übungen werden nach Möglichkeit Datensätze verwendet, die im betriebswirtschaftlichen Kontext interessant sind. So z.B. für die Kundenklassifikation, die Vorhersage von Umsätzen, die Vorhersage von Kundenabwanderung oder die Absatzprognose.

Dieser Kurs richtet sich an Studierende des 2. Semesters, die sich für Verwendung moderner Methoden auf ökonomische Fragestellungen interessieren, und diese im Rahmen von Projekt- und Abschlussarbeiten verwenden wollen. Die unterrichteten Inhalte eröffnen berufliche Perspektiven als Data-Analyst oder Data-Scientist. Der Kurs ist als Folgeveranstaltung zum Kurs *Quantitative Tools 1* geplant.

1 Lernziele

Die Studierenden beherrschen nach dem Kurs die Software Gretl (sowie deren Skriptsprache *Hansl*¹) zur Durchführung komplexer Datenanalysen und können verschiedene maschinelle Lernverfahren selbstständig implementieren und evaluieren. Sie sind in der Lage, fortgeschrittene statistische Methoden wie *Shrinkage Regression*, Regressionsbäume, Random Forests, Support Vector Machines und Clusteralgorithmen auf strukturierte Datensätze anzuwenden und die Ergebnisse professionell zu visualisieren sowie im jeweiligen wirtschaftswissenschaftlichen Kontext zu interpretieren. Darüber hinaus können sie die Vor- und Nachteile verschiedener Methoden kritisch bewerten und ihre Analysen mittels moderner Dashboards präsentieren.

2 Methodik

Die Lehrveranstaltung ist als Vorlesung mit einer begleitenden Übung (insgesamt 4 SWS) konzipiert, wobei die Übungen im PC-Labor stattfinden. Übungsmaterialien und Datensätze werden zur Verfügung gestellt.

3 Kursvoraussetzungen

- Besuch des Kurses Quantitative Tools 1
- Regelmäßige Anwesenheit in Vorlesung und Übung
- Grundlegende Bereitschaft sich Kenntnisse der Ökonometriesoftware Gretl (oder einer anderen Software wie R, Python oder Julia) anzueignen und praktische Übungen damit durchzuführen

¹Mehr Details zu *Hansl* können Sie hier finden: <https://users.wfu.edu/cottrell/hansl.pdf>

Themengebiete

1. Einführung in Data Science (1 SWS)
 - (a) Was ist Data Science?
 - (b) Anforderungsprofil
 - (c) Daten als Grundlage für Wertschöpfung
 - (d) Der Data Science Workflow
 - (e) **Literatur:** Yu and Barter [2024, Kap. 1–2]
2. Data Literacy (2 SWS)
 - (a) Datentypen, -format und -beschaffung
 - (b) Data Literacy und Projekt- und datenmanagement
3. Datenaufbereitung (2 SWS)
 - (a) Datenprüfung
 - (b) Fehlende und fehlerhafte Werte identifizieren
 - (c) Imputationsverfahren
 - (d) Datenformate
 - (e) Regeln für den Umgang mit Datensätzen
 - (f) Pre-processing Schritte: Standardisierung, Enkodierungsverfahren
 - (g) Feature-Building
 - (h) **Literatur:** Yu and Barter [2024, Kap. 4–5]
4. Visualisierung von Daten (2 SWS)
 - (a) Geschichte der Datenvisualisierung
 - (b) Elemente von Grafiken
 - (c) Symbole, Beschriftung, Ausrichtung
 - (d) Typen der Visualisierung
 - (e) **Literatur:** Yu and Barter [2024, Kap. 5], Schwabish [2014], Krause and Rennie
5. CRISP-DM-Standard (0,5 SWS)
 - (a) Standard für die Entwicklung von Data Mining Prozessen
 - (b) **Literatur:** <https://datasolut.com/crisp-dm-standard/>
6. Unsupervised learning (3 Stunden)
 - (a) Dimensionsreduktion mittels PCA und Bi-Plot

- (b) k-means Clustering und Silhouette-Plot
 - (c) **Literatur:** James et al. [2023, Kap. 6.3, 12.1–12.4], Yu and Barter [2024, Kap. 5–7], Greenacre et al. [2022]
7. Regressionsanalyse und out-of-sample Prognostik (**2 SWS**)
- (a) Wiederholung Grundlagen der Regressionsanalyse und KQ-Schätzer
 - (b) Feature-Selection und sparse linear regression Methoden
 - (c) Ridge, Lasso, Elastic-Net
 - (d) **Literatur:** James et al. [2023, Kap. 3.1–3.4], Yu and Barter [2024, Kap. 8–10]
8. Kreuzvalidierung zur Evaluation der Prognosegüte von Modellen (**2 SWS**)
- (a) Training- und Testset
 - (b) Random sampling
 - (c) Evaluationsmetriken
 - (d) **Literatur:** James et al. [2023, Kap. 5], Yu and Barter [2024]
9. Entscheidungsbaum (**2 SWS**)
- (a) Grundlagen von Entscheidungsbäumen
 - (b) Regression-Tree für Regression
 - (c) Vor- und Nachteile von Bäumen
 - (d) **Literatur:** James et al. [2023, Kap. 8.1], Yu and Barter [2024, Kap. 12], Cottrell [2023]
10. Zufallsbäume (**2 SWS**)
- (a) Ensembles
 - (b) Bootstrapping
 - (c) Random Forest
 - (d) Features Importance
 - (e) Gradient boosted trees
 - (f) **Literatur:** James et al. [2023, Kap. 8.2–8.3], Yu and Barter [2024, Kap. 12], Cottrell [2023]
11. Klassifikationsverfahren (**2 SWS**)
- (a) Logit-Modell: binäre Ausprägung
 - (b) Multinominiale Logit-Modelle
 - (c) Grundlagen von Classification-Trees

- (d) Random forest und boosting classification trees für Klassifikation
- (e) Evaluationsmetriken für Klassifikation
- (f) **Literatur:** [James et al. \[2023, Kap. 4, 8.1.2, 8.3.1\]](#), [Yu and Barter \[2024, Kap. 11\]](#), [Cottrell \[2023\]](#)

12. Support Vector Machines (2 SWS)

- (a) Grundlagen von SVM
- (b) **Literatur:** [James et al. \[2023, Kap. 9\]](#), [Yu and Barter \[2024, Kap. 12\]](#), [Cottrell \[2024\]](#)

13. Reporting-Tools und story-telling (2 SWS)

- (a) Relevanz von Reporting-Tools
- (b) Softwarelösungen
- (c) BI-Tools und Software
- (d) Best Practices

Software

Für die statistisch empirische Analyse wird die open-source und kostenfreie Software Gretl benutzt. Diese wird auf dem Labor-PC installiert sein. Sie können diese aber auch auf Ihrem Privatrechner installieren.

Stellen Sie sicher, dass Sie mindestens die Version 2024c verwenden.

Windows: <https://gretl.sourceforge.net/win32>

MAC OS X: <https://gretl.sourceforge.net/osx.html>

LINUX:

- In jeder guten Paketverwaltung erhältlich. Installation in Debian-basierten Systemen mittels Kommandozeile: `sudo apt-get install gretl`
- Für Nerds: Selber kompilieren für Debian bzw. Ubuntu:
https://github.com/atecon/install_gretl_ubuntu

Literatur

Allin Cottrell. Random forest 0.3. 2023. URL https://gretl.sourceforge.net/current_fnfiles/unzipped/random_forest.pdf.

Allin Cottrell. Gretl + SVM. 2024. URL <https://sourceforge.net/projects/gretl/files/manual/gretl-svm.pdf>.

Michael Greenacre, Patrick J. F. Groenen, Trevor Hastie, Alfonso Iodice D'Enza, Angelos Markos, and Elena Tuzhilina. Principal component analysis. *Nature Reviews Methods Primers*, 2(1):1–21, December 2022. ISSN 2662-8449. doi: 10.1038/s43586-022-00184-w. URL <https://www.nature.com/articles/s43586-022-00184-w>.

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An Introduction to Statistical Learning: With Applications in Python*. Springer, Cham, Switzerland, 1st ed. 2023 edition edition, July 2023. ISBN 978-3-031-38746-3. URL https://hastie.su.domains/ISLP/ISLP_website.pdf.download.html.

Andreas Krause and Nicola Rennie. Best Practices for Data Visualisation. URL <https://royal-statistical-society.github.io/datavisguide/>.

Jonathan A. Schwabish. An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives*, 28(1):209–234, February 2014. ISSN 0895-3309. doi: 10.1257/jep.28.1.209. URL <https://www.aeaweb.org/articles?id=10.1257/jep.28.1.209>.

Bin Yu and Rebecca L. Barter. *Veridical Data Science: The Practice of Responsible Data Analysis and Decision Making*. The MIT Press, Cambridge, Massachusetts, October 2024. ISBN 978-0-262-04919-1. URL <https://vdsbook.com/>.

Weitere Materialien

- Video-Lectures zum Buch [James et al. \[2023\]](#): <https://youtube.com/playlist?list=PLoROMvodv4rPP6braWoRt5UCXYZ71GZIQ&feature=shared>
- Weitere Materialien auf der Gretl-project Seite: <https://github.com/gretl-project/material-on-gretl>
- Gretl-Tutorials: <https://github.com/gretl-project/material-on-gretl/wiki>
- Gretl Cheat-Sheet: https://github.com/gretl-project/gretl_cheatsheet/blob/master/datahandling.pdf