



Modul 07-202-2302

Sommersemester 2024

Multivariate Datenanalyse und Data Mining

17.02.2024

Dozent / Übungsleiter:

Roland Schuhr

Umfang / Beginn:

6 SWS / 09.04.2024, 13:15 – 14:45, PC-Pool S-01

Vorlesung:

On Demand E-Learning-Videos.

Die E-Learning-Videos werden durch fünf Q&A-Präsenztermine ergänzt:

Montag, 11:15 – 12:45 (unregelmäßiger Turnus), SR 7 (Institutsgebäude)

Übung:

Computer-Übung.

Dienstag, 13:15 – 14:45, PC-Pool S-01 (Seminargebäude)

Zielgruppe:

Studierende wirtschaftswissenschaftlicher Master-Studiengänge und des Studiengangs M.Sc. Data Science mit Interesse an quantitativer Marketingforschung, Knowledge Discovery in Databases, Business Intelligence sowie empirischer Wirtschaftsforschung.

Inhalt:

Unter Begriffen wie Data Mining, Knowledge Discovery in Databases oder Business Intelligence subsumiert man Strategien, Methoden und Techniken der Exploration wirtschaftlich verwertbarer, zunächst verborgener Information in großen Datenbeständen (Big Data) mit dem Ziel der Entwicklung neuer Geschäftsmöglichkeiten für Unternehmen. Die verwendeten datenanalytischen Methoden haben ihre Wurzeln meist in der Multivariaten Statistik oder in dem zum interdisziplinären Forschungsfeld Artificial Intelligence gehörigen Bereich des Machine Learning.

Die Lehrveranstaltung stellt typische Fragenstellungen des Data Mining und ausgewählte Analysemethoden der Multivariaten Statistik und des Machine Learning vor. Im Vordergrund stehen die methodischen Grundlagen der Dimensionsreduktion/ Feature Extraction sowie der Klassifikation und Diskrimination von empirischen Objekten. In Computer-Übungen werden Datensätze aus Fallbeispielen mit Hilfe der Statistik-Software R analysiert.

Gliederung:

1. Multivariate Daten und Wahrscheinlichkeitsverteilungen
2. Dimensionsreduktion/ Feature Extraction (Hauptkomponentenanalyse)
3. Clusteranalyse (hierarchische Verfahren, k-means, Fuzzy-Verfahren, mclust)
4. Lineare Diskriminanzanalyse und multivariate Varianzanalyse, Nearest Neighbor-Diskrimination
5. Entscheidungsbäume (ID3, C5.0, CART, Bagging und Boosting)
6. Künstliche Neuronale Netze (MLP, Autoencoder, Deep-Learning, SOM)

Universität Leipzig

Wirtschaftswissenschaftliche Fakultät

Institut für Empirische

Wirtschaftsforschung

Grimmaische Straße 12

04109 Leipzig

Telefon

+49 341 97-33 530

Fax

+49 341 97-33 789

E-Mail

schuhr@wifa.uni-leipzig.de

Web

www.wifa.uni-leipzig.de/iew

Postfach intern

171401

Kein Zugang für elektronisch signierte
sowie für verschlüsselte elektronische
Dokumente



Literatur:

- Berk, R.A.: *Statistical Learning from a Regression Perspective*; Springer 2008
- Bishop, C.M.: *Neural Networks for Pattern Recognition*; Oxford University Press 1995
- Burkov, A.: *The hundred-page machine learning book*; Andriy Burkov 2019
<http://themlbook.com/wiki/doku.php>
<http://themlbook.com/wiki/doku.php?id=start&do=index>
<https://freecomputerbooks.com/The-Hundred-Page-Machine-Learning-Book.html>
- Chapman, C. & E. McDonnell Feit: *R for Marketing Research and Analytics*; Springer 2015
- Cios, K.J., Pedrycz, W., Swiniarski, R.W. & L.A. Kurgan: *Data Mining. A Knowledge Discovery Approach*; Springer 2007
- Fahrmeir, L., A. Hamerle & G. Tutz (Hrsg.): *Multivariate statistische Verfahren, 2. Aufl.*; de Gruyter 1996
- Han, J., Kamber, M. & J. Pei: *Data mining. Concepts and techniques, 3rd ed.*; Morgan Kaufmann 2011
- Handl, A. & T. Kuhlenkasper: *Multivariate Analysemethoden. Theorie und Praxis mit R, 3. Aufl.*; Springer 2017
- Hastie, T., Tibshirani, R. & J. Friedman: *The elements of statistical learning. Data mining, inference, and prediction, 2nd ed.*; Springer 2017
- James, G., Witten, D., Hastie, T. & R. Tibshirani: *An Introduction to statistical learning with applications in R, 2nd ed.*; Springer 2021
- Lantz, B.: *Machine Learning with R, 2nd ed.*; Packt Publishing 2015
- Michie, D., D.J. Spiegelhalter & C.C. Taylor (Eds.): *Machine learning, neural and statistical classification*; Ellis Horwood 1994
<http://www1.maths.leeds.ac.uk/~charles/statlog/>
- Mitchell, T.M.: *Machine Learning*; McGraw-Hill 1997 <http://www.cs.cmu.edu/~tom/mlbook.html>

Erwartete Grundkenntnisse:

Gute Kenntnisse in Wahrscheinlichkeitsrechnung, Statistik und linearer Algebra

Modulprüfung:

Klausur (90 Minuten)

moodle-Lernplattform:

Kursteilnehmer können auf Lehrmaterialien und aktuelle Mitteilungen zur Vorlesung und Übung über die Lernplattform Moodle zugreifen.

Der Link des moodle-Kurses wird den Kursteilnehmern zu Beginn des Semesters in AlmaWeb mitgeteilt.