

Modulvorgaben

Modulname: Statistisches Data Mining (StDM)

Code: StDM

Semester u. Klasse: HS 2016

Dozenten: Oliver Dürr

Ziele: Die Studierenden kennen den Data-Mining-Prozess, kennen gängige Verfahren aus dem überwachten und unüberwachten Lernen und können ihre Eignung für vorliegende Daten beurteilen. Sie können die behandelten Methoden mit einem entsprechenden Programm-Paket praktisch umsetzen.

Inhalt: Challenges im Data Mining, Hauptkomponenten-Analyse, Unähnlichkeiten, Multidimensionale Skalierung, gängige hierarchische und partitionelle Clusteranalyse-Methoden, logistische Klassifikation, sowie Klassifikation mit Nächste-Nachbarn-Methode, Regressionsbäume, random forest und boosting, Performance-Evaluation von Klassifikationsmethoden. Statistik-Programm-Paket R: Statistik- und Grafikroutinen für die im Unterricht behandelten Verfahren. Fallbeispiele.

Unterlagen: Folien, Arbeitsblätter, Literatur, Skript

Unterrichtsformen und Notenberechnung: Der Unterricht umfasst 4 Kontaktlektionen Gruppenunterricht pro Woche. 50% der Zeit sind den theoretischen und 50% den praktischen Aspekten des Stoffes gewidmet.

Unterlagen:

Die **Folien**, die während der Vorlesung projiziert werden, stehen zum Download bereit. Somit können diese ausgedruckt werden, um sich Notizen zu machen. Eventuell gibt es nach der Vorlesung noch ein Update der präsentierten Slides

Als Literatur dienen ausgewählte Kapitel des von den Autoren öffentlich zugänglich gemachten Buch „An Introduction to Statistical Learning with Applications in R“ von Gareth James, Daniela Witten, Trevor Hastie und Robert Tibshirani. <http://www-bcf.usc.edu/~gareth/ISL/>

Ausserdem empfohlen werden die zugehörigen Videos des MOOC Kurses dieser Autoren:

<http://www.r-bloggers.com/in-depth-introduction-to-machine-learning-in-15-hours-of-expert-videos/>

Selbststudium:

Wie bereits erwähnt, ist das *Lösen und interaktive Besprechen* der Arbeitsblätter *der wesentliche Teil des Kurses*. Das Studium der Unterrichtsmaterialien und der gegebenen Literatur gehört zum Selbststudium, ersetzt aber nicht den Unterrichtsbesuch. Mit dem Lösen der Übungsblätter kann meistens bereits in der Lektion begonnen werden. Nicht fertiggestellte Aufgaben sind als Hausarbeit zu betrachten!

Das selbständige Bearbeiten der Arbeitsblätter und die Reflektieren der Musterlösungen sind Teil des Selbststudiums. Die Dozenten steht für allfällige Fragen insbesondere während des Praktikums zur Verfügung, Weiter soll der Theorieunterricht anhand der abgegebenen Unterlagen vor- und nachbereitet werden. Dieser im Selbststudium erarbeitete Stoff ist ebenfalls Gegenstand der Modulprüfung.

Bewertung

- Es gibt eine freiwillige Zwischenprüfung. Die Prüfung dauert 45-60 Minuten und ergibt eine erste Vornote, die zu 15% in Endnote zählt, falls sie besser ist als die Note der Schlussprüfung.

- Die freiwillige Bearbeitung einer Hausarbeit ergibt eine zweite Vornote, die zu 10% in Endnote zählt, falls sie besser ist als die Note der Schlussprüfung
- Die Endprüfung ist obligatorisch und dauert 90 Minuten.
- Die Modulendnote ist der grössere Wert von
 $[0.1 \times \text{HA} + 0.15 \times \text{Zwischenprüfung} + 0.75 \times \text{Endprüfung}] (*)$ und
 $[0.15 \times \text{Zwischenprüfung} + 0.85 \times \text{Endprüfung}] (*)$ und
 $[0.1 \times \text{HA} + 0.9 \times \text{Endprüfung}] (*)$ und
[Endprüfung] (*).

Die Vorleistungen (Zwischenprüfung, Praktika) sind insofern fakultative, als sie nur dann zählt, wenn sie die Endnote verbessert. Es wird keine Nachprüfungen für die Vorleistungen geben. Die Endprüfung kann bei begründetem Fernbleiben (Krankheit mit Arztzeugnis, Militärdienst, etc.) nachgeholt werden.

15.09.2016
