

Data Mining und maschinelles Lernen

Johannes Fürnkranz

`juffi@ke.tu-darmstadt.de`

Inhalt

- Einführung in maschinelles Lernen und Data Mining
 - mit Schwerpunkt auf symbolisch/logischen Ansätzen
 - mathematisch/statistisch/symbolische Ansätze werden in Vorlesungen der Kollegen **Jan Peters**, **Stefan Roth**, u.a. behandelt
 - **Kristian Kersting** behandelt Kombinationen von statistischen und logischen Ansätzen
 - Kombinationen sind durchaus sinnvoll

Organisatorisches

- Termine
 - Vorlesung
 - Do 11.40-13.20, S202|C205
 - Übung
 - Di 11.40-13.20, S202|C205
 - Übungsvorbesprechung: 23. 10.
 - VO und UE Termine werden aber gelegentlich getauscht werden
- Benotung:
 - Klausur am Ende des Semesters
 - Teilnahme am praktischen Projekt kann eine Notenverbesserung um 1 Notenstufe bringen
 - Klausur muß aber bestanden werden!

Organisatorisches

- Nächste Woche (25.10., **12.20h**) ein Gastvortrag:

Neural Networks and Art: Case Studies in TensorFlow 25.10.2018

Jonathan Wellons, Software Engineer, Google

**25.10.2018, 12:20 Uhr – 13:05 Uhr | TU Darmstadt, Gebäude S2 | 02, Raum C 205,
Hochschulstraße 10, 64289 Darmstadt**

Veranstalter: [Knowledge Engineering](#)

Referent: Jonathan Wellons, Software Engineer, Google

Beschreibung:

This talk will review the essential elements of Neural Networks and the main differences separating them from other and older ML algorithms.

To illustrate the particular versatility of Neural Networks, I will review a series of projects in the arts that were done on the TensorFlow platform.

They include translation, vision and recognition and computer-generated drawings and music.

Bio:

Jonathan Wellons received his BA from Grinnell College and his Ph.D. in Computer Science from Vanderbilt University. In addition to academia and the private sector, he has worked in government. He has been at Google for 5 years and is currently based in the Zurich office.

Übungen

- Übungen
 - Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung
 - Lösungen werden von Assistenten in den Übungsterminen diskutiert
- Zuständige Assistenten
 - Markus Zopf (zopf@aiphes.tu-darmstadt.de)
 - Sebastian Kauschke (kauschke@ke.tu-darmstadt.de)

Übungen

- Veröffentlichung der Übungsausgabe ist üblicherweise Dienstags auf der Homepage
 - Beziehen sich auf den aktuellen Vorlesungsstoff
- Besprechung der Übung jeweils eine Woche später
 - es wird von den Studierenden erwartet, daß die Übungsaufgaben **vor der Besprechung** selbständig bearbeitet werden
 - Besprechung sollte vor allem dazu dienen, offene Punkte anzusprechen

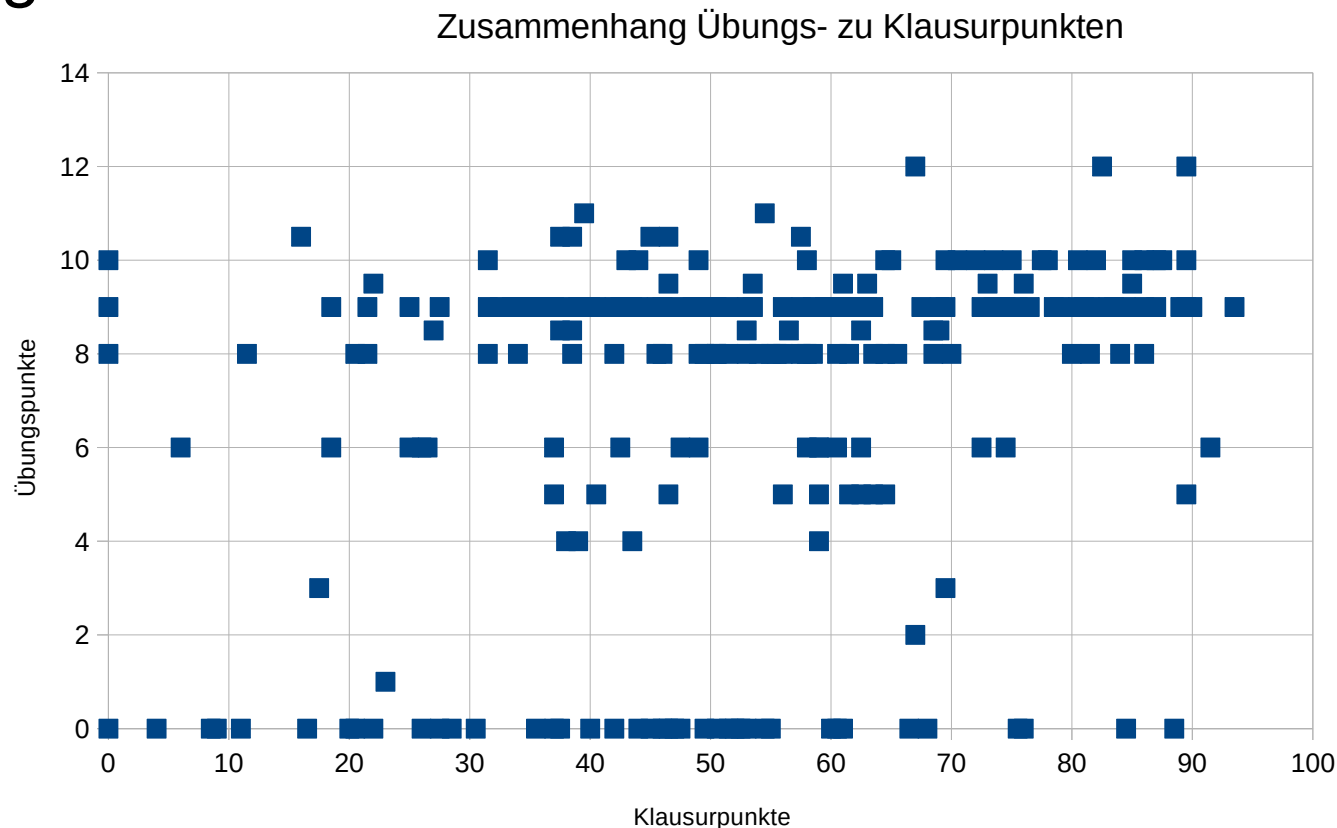
Projekt

- Im Laufe des Semester erhalten Sie ein praktisches Projekt
 - das bis zum Semesterende gelöst werden sollen
 - Lösung in Gruppen (bis zu 3 Teilnehmern) ist erlaubt
- Thematik
 - Praktische Erprobung der vorgestellten Methoden mit Hilfe der Weka-Data Mining Library (open source, Java)
 - Programmierung eines einfachen Algorithmus
- Benotung
 - Erfolgreiche Teilnahme kann eine Verbesserung um bis zu einer Notenstufe führen
 - Nur bei bestandener Klausur

Übung und Projekt

Ziel: Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung

→ Zusammenhang zwischen Teilnahme an Übung und Erfolg in Klausur



Übung und Projekt

Ziel: Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung

→ Zusammenhang zwischen Teilnahme an Übung und Erfolg in Klausur

| Notendurchschnitt und Unterschiede | | | |
|---|---------|---------|---------|
| | WS12/13 | WS13/14 | WS14/15 |
| Endnote mit Übungsteilnahme (inkl. 5,0er) | 2,81 | 2,88 | 3,10 |
| Endnote mit Übungsteilnahme ohne Bonus | +0,48 | +0,34 | +0,35 |
| Endnote ohne Übungsteilnahme | +1,00 | +0,82 | +1,09 |

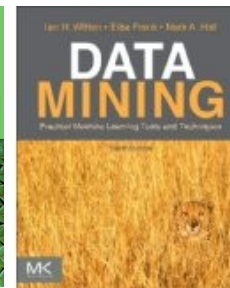
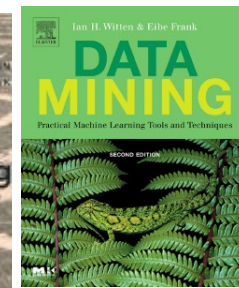
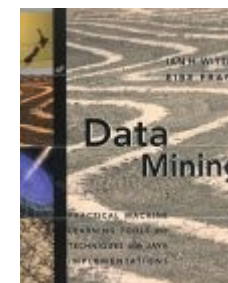
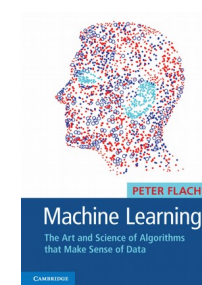
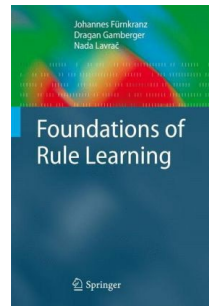
Materialien

- Folien
 - Vorlesungssprache ist Deutsch,
 - but slides will be in English
- Vorlesungs-Homepage
 - <http://www.ke.informatik.tu-darmstadt.de/lehre/ws-18-19/mldm/>
- Moodle wird nur für Projekt genutzt
- Diskussions-Forum
 - <http://www.fachschaft.informatik.tu-darmstadt.de/forum/viewforum.php?f=225>

Bücher

Die folgenden Bücher können als gute Ergänzung für einige der in der Vorlesung gebrachten Kapitel dienen.

- Johannes Fürnkranz, Dragan Gamberger, Nada Lavrac: *Foundations of Rule Learning*, Springer-Verlag 2012
- Tom Mitchell: *Machine Learning*, McGraw-Hill, 1997
- Peter Flach: *Machine Learning*, Cambridge University Press 2012
- Ian Witten & Eibe Frank: *Data Mining*, Morgan Kaufmann, 1999
Deutsche Ausgabe: Hansa, 2001
2nd edition, 2005, 3rd edition 2011



Software

- WEKA: Open Source Java Library
<http://www.cs.waikato.ac.nz/ml/weka/>
- Extensive Documentation available there



Machine Learning Group at the University of Waikato

[Project](#)

[Software](#)

[Book](#)

[Publications](#)

[People](#)

[Related](#)

Weka 3: Data Mining Software in Java

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like **this**, and the bird sounds like **this**.

Weka is open source software issued under the **GNU General Public License**.

We have put together several free online courses that teach machine learning and data mining using Weka. Check out the **website for the courses** for details on when and how to enrol. The videos for the courses are available **on Youtube**.

Yes, it is possible to apply Weka to **big data**!

Weiterführende Vorlesungen

- VO Maschinelles Lernen: Statistische Ansätze (Peters)
 - Schwerpunkt auf neurale Netze, Support-Vektor Machines, ...
- VO Probabilistische Graphische Modelle (Kersting)
- VO Statistical Relational AI (Kersting)
- VO Künstliche Intelligenz
 - Suchen, Planen, Bayes'sche Netze, Neuronale Netze
- VO Deep Learning
 - mehrere Vorlesungen (Methods & Applications, in Vision, in NLP...)
- VO Natural Language Processing & the Web (Gurevych)
 - Mining Wikis etc.
- VO Lernende Roboter (Peters)
 - Reinforcement Learning etc.

Weiterführende Veranstaltungen

- PR Praktikum Maschinelles Lernen und Data Mining
 - Üblicherweise im SS
 - Praktische Anwendung des gelernten
 - Teilnahme an einem Wettbewerb
 - In diesem Semester: **Data Science Praktikum**
- Seminare
 - Finden jedes Semester statt
 - Diskussion aktueller Forschungsarbeiten auf dem Gebiet
- Abschlussarbeiten und Forschungspraktika
 - Kann jederzeit begonnen werden
 - Vorkenntnisse in ML und DM erforderlich
 - Themen (leider nicht ganz aktuell) auf der KE-Seite