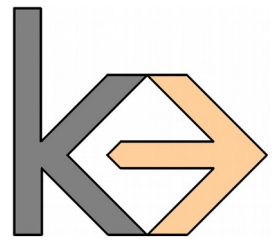




TECHNISCHE
UNIVERSITÄT
DARMSTADT



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: 24. 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!

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

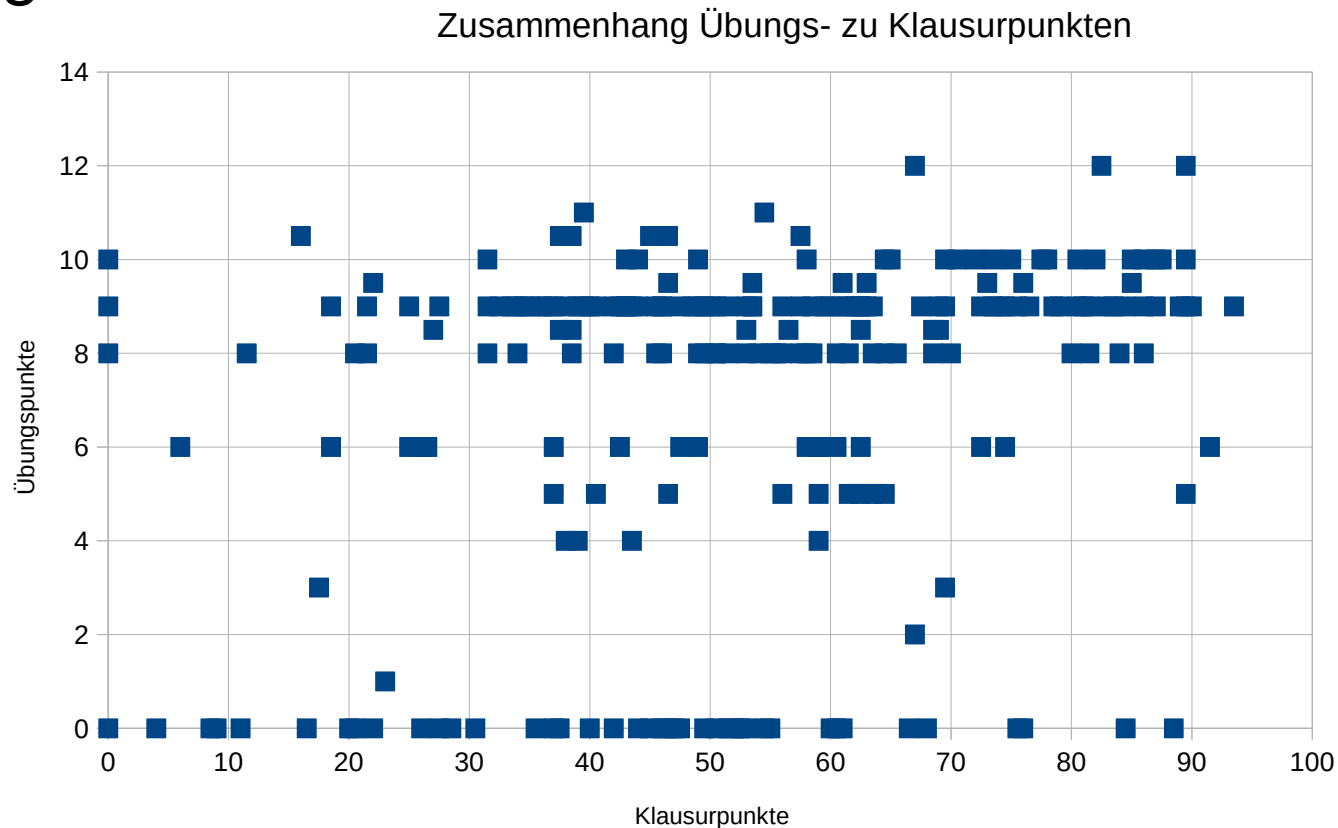
Projekt

- 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

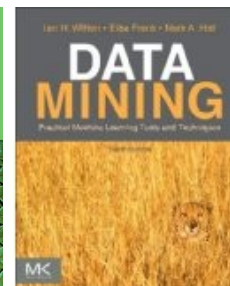
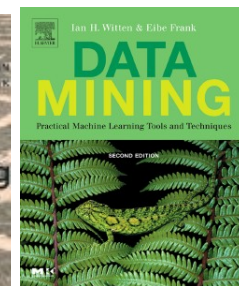
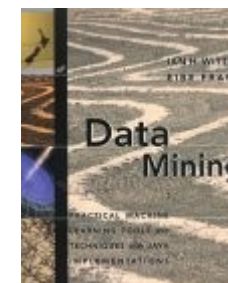
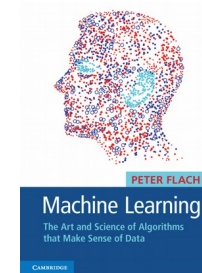
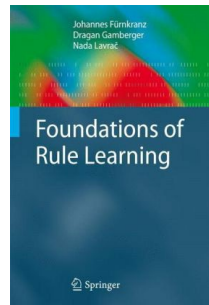
Materialien

- Folien
 - Vorlesungssprache ist Deutsch,
 - but slides will be in English
- Vorlesungs-Homepage
 - `http://www.ke.informatik.tu-darmstadt.de/lehre/ws-17-18/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 Veranstaltungen

- PR Praktikum Maschinelles Lernen und Data Mining
 - Üblicherweise im SS
 - Praktische Anwendung des Gelernten
 - Teilnahme an einem Wettbewerb
- 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