

Prüfungsstoff AIM3

- **Generelles**

- Amdahl's Law, Gustafson's Law
- data selection, data types, data cleansing
- feature extraction
- learning problems (supervised, unsupervised)
- evaluation

- **Map/Reduce, Hadoop**

- Motivation, Anwendungsfälle
- Map/Reduce Paradigma
- Funktionsweise von Hadoop
- HDFS
- Ablauf eines Map/Reduce Vorgangs (Map, Combine, Shuffle, Reduce)
- fortgeschrittene Programmiertechniken (Joins, Secondary Sort)
- Schwächen von Map/Reduce

- **Stratosphere**

- Motivation
- Paradigma, Unterschiede zu Hadoop und Map/Reduce
- Funktionsweise, Operatoren (Cross, Match, Co-Group), Output-Contracts

- **Clustering (Unsupervised Learning)**

- Problembeschreibung, Lösungsansätze
- Anwendungsfälle
- Algorithmus: K-Means
- Parallelisierung

- **Classification (Supervised Learning)**
 - Problembeschreibung, Lösungsansätze
 - Anwendungsfälle
 - Decision Trees
 - Algorithmus: Naive Bayes (skizzieren anhand von Bayes' Theorem)
 - Parallelisierung Naive Bayes skizzieren
- **Collaborative Filtering (Recommendation Mining)**
 - Problembeschreibung
 - Anwendungsfälle
 - Algorithmus: Itembased Collaborative Filtering
 - Parallelisierung skizzieren