Data Mining Lab Course WS 2016/17

Organization and Contents

L. Richter    J.M. Cejuela

Department of Computer Science
Technical University of Munich

Premeeting, Wednesday, Jun 28th
Outline

Contents

Organization
Mission statement: At the end of the course you have gone through the whole path of data mining, from dataset preparation up to meaningful predictions. We provide you the glue to the theory.
Steps

This includes o.a. the following steps:

- Dataset search & Description
- Understanding the data & naïve Introspection
- Feature Construction & Selection
- Prediction & Evaluations
Our tools

- Use scripting languages/scripts for data processing
- Optional: Database
- Use Weka/RapidMiner/Python/R for various processing steps:
  - discretization
  - feature selection
  - learning algorithm implementations
- We will not implement learning schemes
What we provide:

We provide infrastructure and support:

- A lab run in reverse class room style
- A wiki for the documentation of your work and communication with your peers and supervisors
- Computation power: various multi-core machines for the analyses
- Mini-talks to selected topics from theory
- Feedback to your questions and discussion of your findings
What we expect:

- Curiosity
- Self motivation
- Creativity to come up with own solutions
- Question yourself if you prefer clear instructions and directions → this is not your course
- Readiness to retrieve (some) necessary background knowledge autonomously
- **Soft skills**: The ability to form and to collaborate in a team. No lonesome wolves!
Organization

- All of you are asked to suggest interesting datasets
- We select a dataset for each group
- There is one team of 3-4 people for one dataset
- You populate the wiki every week
- You get substantial feedback every week in the plenary meetings
- Grading based on the group’s success