VU Machine Learning

Winter 2014/15

Exercise 1

Exercise 1

- Groups of 2 students
- Perform experiments with regression and classification techniques in machine learning
- Write a report paper
  - around 10 pages, including tables & diagrams
- Submission: 07.12.2014
- Presentations: 10.12 (15:00-19:00)

Exercise 1 – Data Sets

- Pick 2 regression data sets and 2 classification data sets from UCI ML Repository

  http://archive.ics.uci.edu/ml/

  You can also use data sets from other sources
- Must have different characteristics!
  - number of samples – small vs. large
  - number of dimensions – low vs. high dimensional
  - missing values
  - pre-processing needed
  - …
- Choice of diverse data sets important for grading !

Exercise 1 – Techniques

- Chose 2 different regression techniques and 3 different classification techniques
- Argue & justify choice (part of grading…)
- Experiment with different parameter settings
- And report on it !
- Compare results among selected techniques and datasets
- Evaluate effect of pre-processing (e.g. different strategies for missing values, feature scaling, …)

Exercise 1 – Software

- R (http://www.r-project.org/)
- Mathlab (http://www.mathworks.com/discovery/machine-learning.html)
- WEKA (http://www.cs.waikato.ac.nz/ml/weka/)
- Rapid Miner
- Orange
- ..

Bonus points if you use R or Matlab (or other powerful software for regression techniques)

Report

- Report should be around 10-15 pages
- Full report of your work
- Experiments, parameters tried
- Characteristics of data sets & pre-processing (i.e. handling of missing values, scaling etc.)
- Characteristics of regression/classification techniques
- Explanation of choice for data sets & techniques
- Discuss experimental results, compare them in regard of the different datasets & techniques (tables, figures)
- Do not include code in report