VU Machine Learning

Winter 2017/18

Exercise 2

Exercise 2

- Groups of 3 students
- Perform experiments with regression techniques in machine learning
- Prepare a slide presentation
  - around 20-40 slides, including tables & diagrams
- Submission: 19.12.
- Discussion of reports: 20/21.12.

Exercise 2 – Data Sets

- Pick 4 regression data sets
  - 2 from UCI ML Repository
  - 2 from Kaggle (a list of datasets will be provided in TUWEL)
- 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 2 – Techniques

- Chose 3-4 different regression techniques
- Experiment with different parameter settings
- And report on it !
- Compare results among selected techniques and datasets
  - Using different performance metrics for regression methods
- Evaluate effect of pre-processing (e.g. different strategies for missing values, feature scaling, …)

Exercise 2 – Software

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

Presentation

- Around 20-40 slides (report of your work)
- Experiments, parameters tried
- Performance metrics used
- Characteristics of data sets & pre-processing (i.e. handling of missing values, scaling etc.)
- Characteristics of regression 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 presentation