VU Machine Learning

Summer semester 2020

Exercise 1

Exercise 1

- Groups of 3 students

- Perform experiments with regression techniques in machine learning

- Prepare a slide presentation
  - More than 30 slides, including tables & diagrams (no code)
  - No report needed (only if you prefer to write a report)

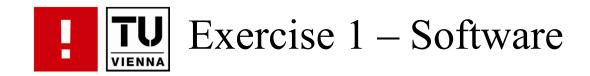- Submission: May 4

- Discussion of reports: After May 4

- Pick 4 regression data sets
  - 1 from exercise 0
  - 1 from UCI ML Repository (or other sources)
  - 2 data sets will be provided in TUWEL

- Must have different characteristics!
  - number of samples – small vs. large
  - number of dimensions – low vs. high dimensional
  - pre-processing needed
  - …

- Chose 4 different regression techniques

- Experiment with different parameter settings

- And report on it !

- Compare results among selected techniques and datasets (4 x 4)

  - Using different performance metrics for regression methods

- Use training and testing set

- Evaluate effect of pre-processing (e.g. feature scaling, …)

- 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)

Discussion of assignment

- Characteristics of data sets & pre-processing (i.e. scaling etc.)

- Characteristics of regression techniques

- Explanation of choice for data sets & techniques

- Experiments, parameters tried

- Evaluation and performance metrics used

- Discuss experimental results, compare them in regard of the different datasets & techniques (tables, figures)

- Discussion of code

- ...