VU Machine Learning

Winter 2017/18

Exercise 3.1

Metalearning (Automated Algorithm Selection)

Exercise 3.1

- Automated selection of machine learning algorithms (Metalearning)

  Based on the Rice's framework (see metalearning slides) generate a model that is able to select automatically the 'best' machine learning algorithm for unseen data sets:

  1. Select several data sets (>15 data sets). Data sets from previous assignments can also be used.
  2. Implement a procedure that extracts from data sets several measurable characteristics/features (> 10 features). Possible features include statistical and model-based features. You can also propose new features.

Exercise 3.1

3. Select 4 - 5 machine learning algorithms (A1, A2, …: default parameters can be used) and apply them on data sets selected in step 1. Record the best performing algorithm (based on classification accuracy) for each data set.
4. Based on steps 1 - 3 generate a new data set that consist of features extracted in step 2 and the class attribute that represents the best performing algorithm (class labels: A1, A2, …). Each instance of this data set describes one data set selected in step 1.
5. Apply 2 - 3 machine learning algorithms on the data set generated in step 4 and compare their performance using cross-validation.

Report

- Report (around 3-4 pages)
  - Describe your framework, features…
  - Report your results

- Presentations/Submission of your framework/report (and the source code) in January. See the information for deadlines/presentation slots in tuwel.

- Groups of 3 students