



Prof. B. Schiele

<schiele AT mis dot tu-darmstadt dot de>

Mario Fritz

<fritz AT mis dot tu-darmstadt dot de>

Übungsblatt 2

Ausgabe: 04.05.2007

Abgabe: **31.05.2007**

Data Mining Cup & Cross-Validation (40 Points)

In this exercise you will run experiments on the data of the DataMiningCup. One of the main focuses of this exercise is to do proper parameter selection by cross-validation in order to assess the performance of your model and prevent over fitting.

- Download data of the DataMiningCup from

<http://www.data-mining-cup.de/2007/Wettbewerb/>

- Import/Convert the data to format you can operate with (e.g. load data into matlab or convert to a particular format). Describe preprocessing steps if you applied any to the data.
- Choose a method appropriate for the problem/data. Explain your choice and write down your expectation in which way you think your approach can be successful.
- Point out how you deal with the cost function specified in the DataMiningCup.
- Identify the parameters of your methods and briefly describe their function.
- Train a model for your data using the method of your choice. Use 5-fold cross validation to prevent over-fitting and to get an estimate of the performance of your system. Report and visualize the mean performance and standard deviation for all the parameter settings you tested. Describe the result of your parameter selection and report the estimated performance.

Presentation of Approach, Results & Lessons Learnt (20 Points)

Prepare a short 10 Minute presentation in which you summarize your approach, results and lessons learnt. The presentation session is scheduled for the 01.06.2007.

Data Mining Cup (10 ExtraPoints)

We award 10 extra points if you submit your results to the DataMiningCup. Therefore you have to register on the website:

<http://www.data-mining-cup.de>

Finally, you do prediction on the test data of the DataMiningCup with the model obtained from the task above. Read the instructions carefully to make a successful submission.

Winning the Challenge (N ExtraPoints)

We award additional points if you perform well in the challenge. If you are among:

- TOP 10 → 60 points
- TOP 25 → 30 points
- TOP 50 → 15 points
- TOP 100 → 5 points

In addition we will have a ranking of all people from this course and award some small prizes.