

Information Mining - winter semester 2019**Exercise sheet 12**

Exercise 1: Combination of multiple models

- (a) In the lecture methods to combine multiple models are shown. Explain briefly the underlying idea behind the following principles: (i) Bagging, (ii) Boosting and (iii) stacking. Done!
- (b) Use *RapidMiner* to classify the example data¹ with the stacking-method. Evaluate the results with a 10-fold cross-validation.
- Use **Naive Bayes** and **k-NN** as learning methods for the models. As the learning method for the stacking use **Decision Tree**.

Exercise 2: k-nn in RapidMiner

In the lecture we have seen that k-nn can be used for instance based classification. An instance is classified by its k neighbors (k is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor. In this exercise use the weather data (knnWeatherDataTraining.arff²) and apply k-nn clustering to classify the following instance. Note, you should consider normalisation before applying k-nn (page 131-132 in chapter 4).

outlook=sunny, temperature=60, humidity=76, windy=TRUE, play=?

Exercise 3: Support-Vector-Machines

Support-Vector-Machines (SVMs) have proven to be a successful method for classifying and are well established.

- (a) Make a sketch of the geometric principle behind the SVMs.
- (b) When using a SVM a over-fitting is unlikely. Why?

Exercise 4: Scheme independent attribute selection

- (a) Do a correlation-based feature selection, like shown in the lecture, with the example data³.
- (b) In RapidMiner you can weight attributes on different criteria. Weight the example data by the following three criteria and evaluate the results.
- (i) correlation

¹http://www.is.inf.uni-due.de/courses/im_ws19/uebung/data_a23.csv

²http://www.is.inf.uni-due.de/courses/im_ws19/uebung/knnWeatherDataTraining.arff

³http://www.is.inf.uni-due.de/courses/im_ws19/uebung/data_a21.csv

- (ii) chi-square
- (iii) uncertainty