

Exercise 5 - Machine Learning II - 20116

Please send your submissions (runnable code, plots and written answers) **via email to weis@ccc.cs.uni-frankfurt.de until Tuesday May 24th 2016**. One submission per student. Prepare to present your solutions in the exercise session. Students that are not able to explain their solutions may not be given credit on their submissions.

1 Regression in General (4 Points)

1. (Textual answer please) Describe in no more than five sentences what regression in general means.
2. (Textual answer please) What is the connection between regression and inference ?
3. (Textual answer please) In what cases would you use logistic regression instead of linear regression ?
4. (Textual answer please) What is the goal of the dual representation in logistic regression ?
5. (Textual answer please) What are the risks in applying regression on a small (probably noisy) dataset, and how could you prevent it ?
6. (Textual answer please) What is the connection between Occams Razor, the curse of dimensionality, and sparse linear regression ?
7. (Textual answer please) What are the benefits of using a Bayesian Model for regression ?
8. (Textual answer please) What is the *kernel trick* ?

2 Logistic regression - application (6 Points)

1. Download wine.data from the files section of the website
2. Suppose that the tests for the wine's attributes are that expensive that we can only chose **2 attributes** to measure for classification, which attributes would you select?
 - Extending the example in http://scikit-learn.org/stable/auto_examples/linear_model/plot_iris_logistic.html, create a logistic regression classifier for any two attributes/features
 - Calculate the accuracy score for each of the classifiers