

# Machine Learning

## Exercise Sheet 4

Prof. Dr. Dr. Lars Schmidt-Thieme, Osman Akcatepe  
Information Systems and Machine Learning Lab (ISMLL)  
University of Hildesheim

29. November 2011  
Deadline 5. December, 14.15

### Section 1: IRLS (5 Points)

Given the following data:

y	x	y	x
0	9.5	1	11.1
0	9.6	1	11.1
0	9.7	1	11.1
0	9.8	1	11.5
0	9.9	1	11.8
0	10.5	1	11.9
0	11.0	1	12.1
0	11.2	1	12.2
0	11.5	1	12.5
0	11.7	1	12.6
0	12.1	1	12.6

**a)**

Use the linear regression (in R) for the target variable  $y$ . Calculate the mean squared error of the model for the given data.

**b)**

Calculate a logistic regression model for the target variable  $y$ . Use the algorithm *iteratively reweighted least squares* (on the paper, you can solve the occurring equation system with calculator). Stop it after the second iteration. Calculate the mean squared error of the model for the given data.

**c)**

Plot the data and the estimated functions from (a) and (b). Discuss the results.

## Exercise 2: Statistical models in R (3 Points)

Read the capital 11 from „An Introduction to R“.

**a)**

With which formula a linear regression would be stated without intercept? Write a formula for a (non-trivial) polynomial regression.

**b)**

How does one make predictions with a model calculated by means of `lm()`?

## Exercise 3: Linear and logistic regression in Weka (2 Points)

Carry out a linear and a logistic regression with a regression and a classification dataset of your choice in Weka.

**a)**

Which of the indicated error measures are meaningful in order to assess a regression and classification model? Name respectively a measure and give reasons shortly.

**b)**

Tinker in both cases with the parameter `ridge` and check out, whether this leads to better results.