Machine Learning
Exercise Sheet 3

Prof. Dr. Dr. Lars Schmidt-Thieme, Martin Wistuba
Information Systems and Machine Learning Lab
University of Hildesheim

November 10th, 2015
Submission until November 17th, 13.00 to wistuba@ismll.de

Exercise 5: Discriminant Analysis (5 Points)

Scientists compared the earth of Iowa which contains a specific bacterium (class 1) with other earth that does not contain it (class 2). They observed the variables x_1 (pH value) and x_2 (nitrogen content). The number of instances pro class, the mean of the vectors and the covariance matrix for both kind of earths is given as follows:

$$n_1 = 13, \qquad n_2 = 10$$
$$\mu_1 = \begin{pmatrix} 7.8 \\ 43 \end{pmatrix}, \qquad \mu_2 = \begin{pmatrix} 5.9 \\ 18.8 \end{pmatrix}$$
$$\Sigma_1 = \begin{pmatrix} 0.5 & 6 \\ 6 & 140.2 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} 0.1 & 0.17 \\ 0.17 & 20.2 \end{pmatrix}$$

a) Estimate the discriminant functions for both classes.

b) Assign the observation $x = \begin{pmatrix} 6 & 52.5 \end{pmatrix}^T$ to one of the both classes.

c) Is this a linear or a quadratic discriminant analysis? Mention differences between LDA and QDA.

Exercise 6: Regularization / Weka Grid Search (5 Points)

a) What is meant by the term overfitting and how it comes to pass?

b) How can you recognize that a model is overfitted?

c) How can you prevent overfitting?

d) WEKA

- Install Weka: `http://www.cs.waikato.ac.nz/ml/weka/`.
- Save the data for Weka from `http://repository.seasr.org/Datasets/UCI/arff/spect_train.arff`.

- In the Weka-Explorer open your data-file and in the next step choose Logistic *Classify→Choose→functions→Logistic*
- Define a grid and apply a grid search for the `ridge` parameter. Plot the results.