Machine Learning
Exercise Sheet 2

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

October 28th, 2014
Submission until November 4th, 13.00 to wistuba@ismll.de

Exercise 5: Gradient Descent (5 Points)

Apply gradient descent on the function $f(x) = \frac{1}{4}x^4 + \frac{1}{3}x^3 - \frac{1}{2}x^2$ with following configurations:

a) Use step length $\alpha = 0.3$ and starting point $x_0 = -1$ and show the first four iterations. What is your minimum?

b) Use step length $\alpha = 2$ and starting point $x_0 = -1$ and show the first four iterations. What has happened and why?

c) Use step length $\alpha = 0.3$ and starting point $x_0 = 0$ and show the first two iterations. What has happened and why?
Do the same again with $\alpha = 0.8$ and starting point $x_0 = 0.5$ and show the first four iterations. Where is your minimum now?
What would be a possible solution to overcome the problem just identified?

Exercise 6: Logistic Regression (5 Points)

Given is 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) Apply linear regression for the target y using the method of R (see Exercise 2). Estimate the mean squared error of the model for the given data. Submit results and the code.

b) Optimize a logistic regression model with the Newton algorithm for the target y. Use $\alpha = 1$ and $\beta = 0$ for this and do it manually (for solving equation systems you can use a tool e.g. R). Stop after two iterations and estimate the mean squared error of the model on the given data. Compare the parameters after the first and second iteration with the final parameters $\hat{\beta}_0 = -23.35$, $\hat{\beta}_1 = 2.064$

c) Plot the data and the predictor functions of (a) and (b). Discuss the results.

Exercise 7: 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 8: 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.