Machine Learning
Exercise Sheet 1

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

October 24th, 2014
Submission until October 31st, 13.00 to schilling@ismll.de or baker@ismll.de

Exercise 1: Linear Regression & Error Measures (10 Points)

a) Given are the data instances of the example from the lecture (gas consumption):

$$\mathcal{D} = \{(2, 6), (6, 5), (8, 4.5)\}$$

Estimate the target $\hat{y}$ for $x = 10$ using the method of least squares. The true value is $y = 2$. Estimate the error. Interprete the result. Create a plot of all distances and show for each data point the least square error.

b) A website collects DVD ratings and then uses them to recommend users a DVD. Given are the ratings of two users among all items (1 Star is the worst rating, 5 the best):

Index	**User**	**DVD**	**Rating**
1	A	*The Big Lebowski*	4 Stars
2	A	*Brazil*	2 Stars
3	A	*Titanic*	5 Stars
4	B	*Brazil*	3 Stars
5	B	*The Godfather*	4 Stars
6	B	*Toy Story*	4 Stars

Three different regression models $\hat{r}_1, \hat{r}_2, \hat{r}_3$ are learned and make the following predictions:

Index	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$
1	3.7	3.8	3.9
2	2.4	2.5	2.3
3	2.2	3.0	4.1
4	3.2	3.1	2.9
5	4.7	4.4	4.2
6	4.1	3.9	4.2

Estimate for every regression model the mean squared error and the mean absolute error, which is given by

$$\text{MAE}(\hat{y}, \mathcal{D}) = \frac{1}{N} \sum_{n=1}^{N} |y_n - \hat{y}(x_n)|$$

in comparison to the true ratings. What are the differences between both error measures?

Exercise 2: Optimality of Linear Regression Parameters (10 Points)

In the lecture it was shown that for the simple linear regression

$$\hat{\beta}_1 = \frac{\sum_{n=1}^{N}(x_n - \bar{x})(y_n - \bar{y})}{\sum_{n=1}^{N}(x_n - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

minimize the residual sums of squares (RSS).
Proove these equations by computing the partial derivatives of the loss function.
Justify that the given solutions are indeed a global minimizer for the loss.