Machine Learning 1
Exercise Sheet 1

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond, Maurício Camargo
Information Systems and Machine Learning Lab
University of Hildesheim

26/10/2017
Submit your pdf via learn-web

Questions (NOT YOUR SHEET WITH ANSWERS) should be mailed to:

radrumond@ismll.de OR mauriciogmc@gmail.com
(use "[ML1] Exercise Sheet 1" on the topic please)
deadline: 3/11/2017 12:00 (noon)

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

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

$$\mathcal{D} = \{(1, 6),(5, 5),(7, 4.5)\}$$

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

b) A car-rental company collected feedback from their clients about their experience with each rented vehicle in order to recommend them to other users. Given are the ratings of two users among all items (1 Star is the worst rating, 5 the best):

Index	User	Car	Rating
1	A	Renault LeCar	3
2	A	AMC Pacer	1
3	A	Ford Gran Torino	4
4	B	Ford Econoline	3
5	B	Beetle 2018VW	3
6	B	Ford Model T	5

Three different regression $\hat{r}_1, \hat{r}_2, \hat{r}_3$ models were learned and were able to peform the following predictions:

Index	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$
1	3.7	3.8	2.9
2	1.4	1.5	1.3
3	2.2	3.0	4.1
4	3.2	3.1	2.9
5	4.1	3.8	3.5
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:

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

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

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

$$\hat{\beta} = \bar{y} - \hat{\beta}\bar{x}$$

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