Machine Learning
Exercise Sheet 5

Prof. Dr. Dr. Lars Schmidt-Thieme, Rafael Rego Drumond
HiWi's: Khouloud Sallami, Maurício Camargo, Shayan Jawed, Sami Diaf
Information Systems and Machine Learning Lab
University of Hildesheim

November 23th, 2017
Submission until November 30th (noon) by dropping at MACHINE LEARNING postbox
(please indicate in which tutorial are you participating!)
The postboxes is located inside the Samelsonplatz building to the right.

Exercise 9: Hypothesis Tests for LR Parameters (10 Points)

For the following data:

x_1	x_2	y
1	2	2
4	-2	-3.3
2	3	3.5
-5	-1	-2.8
-1	2	1.8

a linear regression model given by the parameters

$$\beta^\top = \begin{pmatrix} -1 & 0.3 & 1.4 \end{pmatrix}$$

has been learned. Perform a (two sided) hypothesis test on all of the parameters, where

$$H_0 : \beta_i = 0 \qquad H_1 : \beta_i \neq 0$$

in order to determine which parameters are significant. What are the resulting standardized coefficients? Use a significance level of $\alpha = 0.05$. Which variables are significant?

Exercise 10: (Stochastic) Coordinate Descent for Polynomial Regression (10 Points)

a) Describe the main difference between regular gradient descent and coordinate descent!

b) Stochastic Gradient Descent (SGD) works very similar to normal gradient descent, the key difference is that only one data instance is used per update, i.e. the (regression) loss function for a single instance resolves to:

$$\mathcal{L}(y, \hat{y}(x)) = \frac{1}{2}(\hat{y}(x) - y)^2$$

For a polynomial regression of order two, i.e.

$$\hat{y}(x) = \beta_0 + \sum_{i=1}^{p} \beta_i x_i + \sum_{l=1}^{p} \sum_{j=l}^{p} \beta_{lj} x_l x_j$$

Compute the update equations of a stochastic coordinate descent for all parameters β_0,β_i and β_{lj} for the single instance loss!