ML 2017/2018: Graded Assignment 1 (5 points)

Deadline: October 10

1. **3 points** This question is about *vectorization*, i.e. writing expressions in matrix-vector form. The goal is to vectorize the update rule for multivariate linear regression.

   (a) Let ϑ be the parameter vector $\vartheta = (\theta_0\, \theta_1 \cdots \theta_n)^T$ and let the i-th data vector be: $x^{(i)} = (x_0\, x_1 \cdots x_n)^T$ where $x_0 = 1$. What is the vectorial expression for the hypothesis function $h_\vartheta(x)$?

   (b) What is the vectorized expression for the cost function: $J(\vartheta)$ (still using the explicit summation over all training examples).

   (c) What is the vectorized expression for the gradient of the cost function, i.e. what is:

   $$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial J(\boldsymbol{\theta})}{\partial \theta_0} \\ \vdots \\ \frac{\partial J(\boldsymbol{\theta})}{\partial \theta_n} \end{pmatrix} \tag{1}$$

   Again the explicit summation over the data vectors from the learning set is allowed here.

   (d) What is the vectorized expression for the ϑ update rule in the gradient descent procedure?

   (e) (bonus points) Vectorization can be taken one step further. We can remove the explicit summation over the training samples by 'hiding´ it in a matrix vector multiplication. Start by collecting all training samples in a data matrix X such that every *row* of X is a vector from the training set (with the augmented $x_0 = 1$ elements, i.e. the first column of X has elements equal to 1).

2. **2 points** Derive an equation that can be used to find the optimal value of the parameter θ_1 for univariate linear regression without doing gradient descent. This can be done by setting the value of the derivative equal to 0. You may assume that the value of θ_0 is fixed.