Assignment 1

Computational Intelligence, SS2017

Team Members		
Last name	First name	Matriculation Number
Kopf	Christian	1331187
–	–	–

I. Task 1

A. 1.1 Derivation of Regularized Linear Regression

The task is to show that

$$\boldsymbol{\theta}^* = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y} \tag{1}$$

is the analytical solution for the optimal parameters to minimize the linear regression cost of the regularized cost function

$$J(\boldsymbol{\theta}) = \frac{1}{m}||\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}||^2 + \frac{\lambda}{m}||\boldsymbol{\theta}||^2 \quad . \tag{2}$$

The minimum of the cost function can be determined by setting the gradient of the cost function to 0.

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}^T \tag{3}$$

With this in mind we calculate the derivative of the regularized cost function.

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{m}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^T \frac{\partial(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})}{\partial \boldsymbol{\theta}} + \frac{2\lambda}{m}\boldsymbol{\theta}^T \tag{4}$$

This leaves us with a solution including the inner derivation which can again be solved to get

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{2}{m}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^T\boldsymbol{X} + \frac{2\lambda}{m}\boldsymbol{\theta}^T \quad . \tag{5}$$

From (5) and the auxilary condition (3) we can set

$$\frac{2}{m}(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^T\boldsymbol{X} + \frac{2\lambda}{m}\boldsymbol{\theta}^T = \mathbf{0}^T \quad . \tag{6}$$

We also may neglect the factor $2m^{-1}$ since it drops out when multiplied by 0.

$$(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^T\boldsymbol{X} + \lambda\boldsymbol{\theta}^T = \mathbf{0}^T \tag{7}$$

By transposing we get

$$\boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y}) + \lambda\boldsymbol{\theta} = \mathbf{0} \quad . \tag{8}$$

Further resolving leads to

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{X}^T\boldsymbol{y} + \lambda\boldsymbol{\theta} = \mathbf{0} \tag{9}$$

$$\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\theta} + \lambda\boldsymbol{\theta} = \boldsymbol{X}^T\boldsymbol{y} \tag{10}$$

$$(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})\boldsymbol{\theta} = \boldsymbol{X}^T\boldsymbol{y} \tag{11}$$

$$\boldsymbol{\theta} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y} = \boldsymbol{\theta}^* \tag{12}$$

which corresponds to (1). The optimal parameters $\boldsymbol{\theta}^*$ only exist if the inverse of $\boldsymbol{X}^T\boldsymbol{X} + \lambda$ exists.

II. Task 2