Assignment 2

Computational Intelligence, SS2020

Team Members		
Last name	First name	Matriculation Number
Blöcher	Christian	01573246
Bürgener	Max	01531577

1 Linear regression

1.1 Derivation of Regularized Linear Regression

- Why is the design matrix $\boldsymbol{X}$ containing $n+1$ and not just n?
  - The simplest form of linear regression is calculated by:

$$y(\boldsymbol{x}, \boldsymbol{\omega}) = \underbrace{\omega_0}_{bias} + \boldsymbol{\omega}^T \boldsymbol{w}$$

  - The additional column is used to compensate the sum of the bias ω_0 and thus to simplify our calculation.

$$\boldsymbol{X} = \begin{bmatrix} 1 & \dots & x_1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & x_N \end{bmatrix} \begin{bmatrix} \omega_0 \\ \vdots \\ \omega_N \end{bmatrix}$$

- What is the dimension of the gradient vector?
  - The gradient contains the derivation of all $J(\boldsymbol{\theta})$ for all variables. That means the gradient has the dimension $m \times 1$ and it contains a group of targets.

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial J(\boldsymbol{\theta})}{\partial \theta_1} \\ \vdots \\ \frac{\partial J(\boldsymbol{\theta})}{\partial \theta_m} \end{bmatrix}$$

- What is the definition of the Jacobian matrix and what is the difference between the gradient and the Jacobian matrix?
  - The Jacobian matrix elements are calculated by the derivatives of all outputs of a vector with respect to all parameters.

$$\boldsymbol{J} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

  - The gradient is computed from a scalar. It is a vector which contains the derivatives of a function for all parameters.

- What is the dimension of the Jacobian matrix and what is it equal to?
  - The dimension of the Jacobian matrix equals the dimension of the Design matrix

- Minimization of the regularized linear regression cost function

$$\begin{aligned} J(\boldsymbol{\theta}) &= \tfrac{1}{m}\|\boldsymbol{X\theta} - \boldsymbol{y}\|^2 + \tfrac{\lambda}{m}\|\boldsymbol{\theta}^2\| \\ J(\boldsymbol{\theta}) &= \tfrac{1}{m}\left[(\boldsymbol{X\theta} - \boldsymbol{y})^T \cdot (\boldsymbol{X\theta} - \boldsymbol{y})\right] + \tfrac{\lambda}{m}(\boldsymbol{\theta^T\theta}) \quad \mid \tfrac{\partial}{\partial \boldsymbol{\theta}} \end{aligned}$$

Using Hint 2 from our exercise sheet the calculation is simplified to:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \tfrac{2}{m}(\boldsymbol{X\theta} - \boldsymbol{y})^T \cdot \boldsymbol{X} + \tfrac{2\lambda}{m}\, \boldsymbol{\theta}^T \quad |\stackrel{!}{=} 0$$

We set the gradient to zero to calculate our solution θ

$$
\begin{aligned}
\tfrac{2}{m}(\boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X} - \boldsymbol{y}^T\boldsymbol{X}) + \tfrac{2\lambda}{m}\boldsymbol{\theta}^T &= 0 \\
\boldsymbol{\theta}^T\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{\theta}^T &= \boldsymbol{y}^T\boldsymbol{X} \\
\boldsymbol{\theta}^T(\boldsymbol{X}^T\boldsymbol{X} + \lambda E) &= \boldsymbol{y}^T\boldsymbol{X} \\
\boldsymbol{\theta}^T &= (\boldsymbol{y}^T\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{E})^{-1} \quad | \; ()^T
\end{aligned}
$$

Due to the property of transpose: $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$ we get:

$$
\begin{aligned}
\theta &= (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{E})^{-1^T}(\boldsymbol{y}^T\boldsymbol{X})^T \\
\theta &= (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{E})^{-1}(\boldsymbol{X}^T\boldsymbol{y})
\end{aligned}
$$

1.2 Linear Regression with polynomial features

1.3 (Bonus) Linear Regression with radial basis functions

2 Logistic Regression

2.1 Derivation of Gradient

2.2 Logistic Regression training with gradient descent