Ludwig-Maximilians-Universitaet Muenchen
Institute for Informatics
Prof. Dr. Volker Tresp
Rajat Koner
Andreas Lohrer
Sebastian Schmoll

Machine Learning
Summer 2021
Exercise Sheet 3

Exercise 3-1 Linear Regression

Let $\mathbf{X} \in \mathbb{R}^{N \times D}$ be a dataset with N samples of dimension D in which the first column contains only ones (to represent the bias), $\mathbf{y} \in \mathbb{R}^N$ a vector with the target values, and $\mathbf{w} \in \mathbb{R}^D$ the weight vector we want to learn. Given the scalar loss

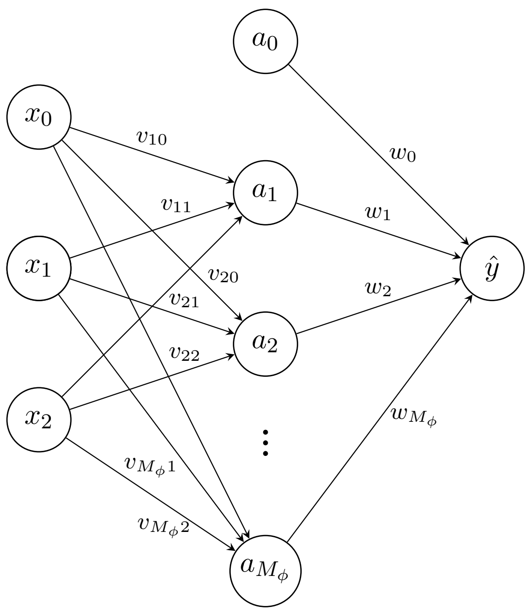
$$L = (\mathbf{y} - \mathbf{X}\mathbf{w})^T(\mathbf{y} - \mathbf{X}\mathbf{w}) + \lambda \mathbf{w}^T\mathbf{w},$$

show that the analytical solution that minimizes the loss L is $\mathbf{w}^* = (\mathbf{X}^T\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^T\mathbf{y}$.

Hint: Use the following identities: $(\mathbf{U}\mathbf{V})^T = \mathbf{V}^T\mathbf{U}^T$, $\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$, $\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}^T$ and $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$.

Exercise 3-2 A simple Neural Network

The illustration below depicts a two-layered neural network with inputs $x \in \mathbb{R}^2$ and for each input one bias term $x_0 = 1$, i.e. $\mathbf{x}_i = (1, x_{i,1}, x_{i,2})^T$. Analogously, there is a bias $a_0 = 1$ in the hidden layer.



As activation function for the hidden neurons we employ a sigmoid, i.e.

$$a_h = \sigma(\mathbf{x}_i^T\mathbf{v}_h) = \frac{1}{1 + e^{-\mathbf{x}_i^T\mathbf{v}_h}}; \qquad \forall\, h = 1, \ldots, M_\phi.$$

The output value $\hat{y}$ is calculated as weighted sum of the neurons in the hidden layer: $\hat{y} = \mathbf{a}^T\mathbf{w} = \sum_{j=0}^{M_\phi} a_j w_j$.

(a) Prove that the following holds: $\frac{\partial a_h}{\partial \mathbf{v}_h} = a_h \left(1 - a_h\right) \mathbf{x}_i$

(b) Express the maximal value of $\hat{y}$ in terms of $\mathbf{w}$ if all weights w_h $(h \in \{0, \ldots, M_\phi\})$ are positive. What's the minimal value?

(c) If $v_{h,j} = 0$ for all $j \in \{0, \ldots, M\}, h \in \{1, \ldots, M_\phi\}$, then what is $\hat{y}$?

Exercise 3-3 PyTorch - Feed Forward Neural Network

On the course website you will find an ipython notebook leading you through the implementaion of a simple feed forward neural network in PyTorch for classifying handwritten digits from the MNIST dataset.