Exercise

10

TUM Department of Informatics

Supervised by Prof. Dr. Stephan Günnemann

Informatics 3 - Professorship of Data Mining and Analytics

Submitted by Marcel Bruckner (03674122)

Julian Hohenadel (03673879)

Kevin Bein (03707775)

Submission date Munich, January 2, 2020

Deep Learning II

Problem 1:

a) Since all weights are positive, $\max(0, {W_j^*}^T x_i) = {W_j^*}^T x_i$ and the following holds:

$$\begin{aligned}
\mathcal{L}_{NN}(W_{NN}^*) &= \frac{1}{2}\sum_{i=1}^{N}(-y_i)^2 \\
&= \frac{1}{2}\sum_{i=1}^{N}({W_{L+1}^*}^T({W_L^*}^T \max(0, \dots(0, {W_2^*}^T \max(0, {W_1^*}^T x_i)))) - y_i)^2 \\
&= \frac{1}{2}\sum_{i=1}^{N}({W_{L+1}^*}^T({W_L^*}^T(\dots({W_2^*}^T({W_1^*}^T x_i)))) - y_i)^2 \\
&= \frac{1}{2}\sum_{i=1}^{N}({W_{L+1}^*}^T \cdot \dots \cdot {W_1^*}^T x_i - y_i)^2 \\
&= \frac{1}{2}\sum_{i=1}^{N}({W_{NN}^*}^T x_i - y_i)^2
\end{aligned}$$

Since the solutions W_{NN}^* and w_{LS}^* are a global optimum, we can set $W_{NN}^* = w_{LS}^*$ and we get $\mathcal{L}_{NN}(W_{NN}^*) = \mathcal{L}_{LS}(w_{LS}^*)$.

b) Since simple linear regression is a special case of a Feed-Forward Neural Network, the Loss is the same. This does not work the other way around because NNs can learn more complex functions. w_{LS}^* being non-negative does not imply anything about the optimal weights of the network W_{NN}^*.Therefore we can conclude that $\mathcal{L}_{NN}(W_{NN}^*) \leq \mathcal{L}_{LS}(w_{LS}^*)$.

Problem 2:

Appendix

We confirm that the submitted solution is original work and was written by us without further assistance. Appropriate credit has been given where reference has been made to the work of others.

Munich, January 2, 2020, Signature Marcel Bruckner (03674122)

Munich, January 2, 2020, Signature Julian Hohenadel (03673879)

Munich, January 2, 2020, Signature Kevin Bein (03707775)