Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 9

Exercise 1: Positive Homogeneity of a Deep Rectifier Network (20 P)

A function $f(\boldsymbol{x})$ is positive homogeneous if $\forall_{t\geq 0}: \ f(t\boldsymbol{x}) = tf(\boldsymbol{x})$. We consider a deep rectifier neural network. The network takes some input $\boldsymbol{x} \in \mathbb{R}^d$. The first layer of activation is defined by the input itself:

$$\boldsymbol{a}^{(0)}(\boldsymbol{x}) = \boldsymbol{x}$$

Then, each pair of consecutive layers of activation is related through the equation:

$$a_k^{(l+1)}(\boldsymbol{x}) = \max\left(0\,,\, \textstyle\sum_j a_j^{(l)}(\boldsymbol{x}) \cdot w_{jk}^{(l)}\right)$$

finally, the top-layer is given by $\boldsymbol{a}^{(L)}(\boldsymbol{x})$. We would like to show positive homogeneity of the top layer by induction. The first layer $\boldsymbol{a}^{(0)}(\boldsymbol{x})$ is trivially positive homogeneous with $\boldsymbol{x}$. In the following question, we consider the induction step.

(a) *Show* that if $\forall_j: \ a_j^{(l)}(\boldsymbol{x})$ is first-order positive homogeneous with $\boldsymbol{x}$, then $a_k^{(l+1)}(\boldsymbol{x})$ is also first-order positive homogeneous with $\boldsymbol{x}$.

Exercise 2: Taylor Decomposition (20 P)

Consider the simple radial basis function

$$f(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{\mu}\| - 1$$

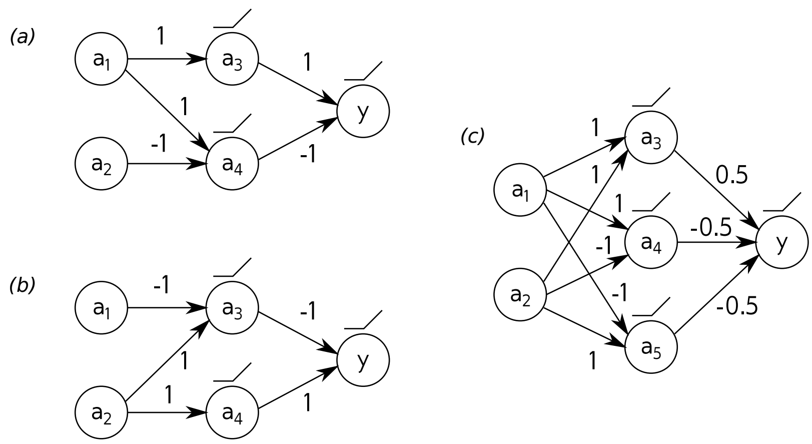
For the purpose of extracting an explanation, we would like to build a first-order Taylor expansion of the function at some root point $\widetilde{\boldsymbol{x}}$. We choose this root point to be taken on the segment connecting $\boldsymbol{\mu}$ and $\boldsymbol{x}$ (we assume that $f(\boldsymbol{x}) > 0$ so that there is always a root point on this segment).

(a) *Show* that the first-order terms of the Taylor expansion are given by

$$R_i = \frac{(x_i - \mu_i)^2}{\|\boldsymbol{x} - \boldsymbol{\mu}\|^2} \cdot (\|\boldsymbol{x} - \boldsymbol{\mu}\| - 1)$$

Exercise 3: Layer-Wise Relevance Propagation (30 P)

We would like to test the dependence of layer-wise relevance propagation (LRP) on the structure of the neural network. For this, we consider the function $y = \min(a_1, a_2)$, where $a_1, a_2 \in \mathbb{R}^+$ are the input activations. This function can be implemented as a ReLU network in multiple ways. Three examples are given below.



We consider the propagation rule:

$$R_j = \sum_k \frac{a_j w_{jk}^+}{\sum_j a_j w_{jk}^+} R_k$$

where $()^+$ denotes the positive part.

(a) *Give* for each network an analytic solution for the scores R_1 and R_2 obtained by application this propagation rule at each layer. More specifically, express R_1 and R_2 as a function of the input activations a_1 and a_2.

Exercise 4: Programming (30 P)

Download the programming files on ISIS and follow the instructions.