Exercises for the course
Deep Learning 2
Summer Semester 2023

Machine Learning Group
Faculty IV – Electrical Engineering and Computer Science
Technische Universität Berlin

Exercise Sheet 9 - Bonus

Exercise 1: Analysis of a similarity models (0 P)

We consider here similarity models of type $y(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x}') \rangle$ with the dot product on a feature map $\phi \colon \mathbb{R}^d \to \mathbb{R}^h$ and satisfying first-order positive homogeneity i.e. $\forall_{\boldsymbol{x}}, \forall_{t>0} : \phi(t\boldsymbol{x}) = t\phi(t\boldsymbol{x})$.
In the following we focus on Linear/ReLU layers:

$$\begin{aligned} a_k &= \big(\textstyle\sum_j a_j w_{jk}\big)^+ \\ a_{k'} &= \big(\textstyle\sum_{j'} a_{j'} w_{j'k'}\big)^+, \end{aligned}$$

with activations a_j and weights w_{jk} and $(\cdot)^+$ indicating the ReLU function. Further assume root points $(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{x}}') = (\varepsilon\widetilde{\boldsymbol{x}}, \varepsilon\widetilde{\boldsymbol{x}}')$ with ε almost zero.

(a) Write down the Taylor expansion of function $y(\boldsymbol{x}, \boldsymbol{x}')$ up to second-order terms.

(b) Analyse zero-order terms. Why do they vanish?
Now, assume the following propagation rule for the Linear/ReLU layer to identify relevant interaction between a pair of neurons j and j':

$$\begin{aligned} R_{jj'} &= \textstyle\sum_{kk'} R_{jj' \leftarrow kk'} \\ &= \sum_{kk'} \frac{a_j a_{j'} \rho(w_{jk}) \rho(w_{j'k'})}{\sum_{jj'} a_j a_{j'} \rho(w_{jk}) \rho(w_{j'k'})} R_{kk'} \end{aligned}$$

(c) Show that $R_{jj'}$ factorizes as $R_{jj'} = \sum_{m=1}^h R_{jm} R_{j'm}$. Use the factorization of the subsequent layer $R_{kk'} = \sum_{m=1}^h R_{km} \cdot R_{k'm}$.