Exercises for the course
Deep Learning 1
Winter Semester 2024/25

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

Exercise Sheet 12

Exercise 1: Building Neural Networks (20 + 20 P)

We consider the problem of learning decision functions in $\mathbb{R}^2$ where $x = (x_1, x_2)$ denotes the two-dimensional input vector. For this exercise, you only have access to neurons of the type

$$a_j = \sigma\Big(b_j + \sum_i a_i w_{ij}\Big)$$

where σ is the step function, i.e.

$$\sigma(t) = \begin{cases} 1 & \text{if } \; t > 0 \\ 0 & \text{if } \; t \leq 0 \end{cases}$$

(a) Construct a neural network that implements the decision boundary below

$$y = \begin{cases} 1 & \text{if } \; \max(x_1, x_2) > 2 \\ 0 & \text{if } \; \max(x_1, x_2) < 2 \end{cases}$$

Specifically, *draw* the neural network graph, and *indicate* for each neuron its weights and bias. (The exact behavior at the decision boundary does not need to be enforced.)

(b) Repeat the exercise for the decision function

$$y = \begin{cases} 1 & \text{if } \; \|x\|_1 > 2 \\ 0 & \text{if } \; \|x\|_1 < 2 \end{cases}$$

Exercise 2: Condition Number (20 + 10 P)

Consider a supervised dataset composed of inputs $\boldsymbol{x}_1, \dots, \boldsymbol{x}_N \in \mathbb{R}^d$ and respective targets $t_1, \dots, t_N \in \mathbb{R}$. Assume that $\frac{1}{N}\sum_{i=1}^N \boldsymbol{x}_i = \boldsymbol{0}$, i.e. the data is centered. Consider the homogeneous linear model $y = \boldsymbol{w}^\top \boldsymbol{x}$, with $\boldsymbol{w} \in \mathbb{R}^d$ a vector of parameters to be learned, and the regularized mean square objective:

$$\mathcal{E}(\boldsymbol{w}) = \alpha \|\boldsymbol{w}\|^2 + \frac{1}{N}\sum_{i=1}^N (\boldsymbol{w}^\top \boldsymbol{x}_i - t)^2$$

we would like to minimize.

(a) *Show* that the Hessian of the error function $\mathcal{E}(\boldsymbol{w})$ is given by the constant matrix:

$$H(\boldsymbol{w}) = 2(\Sigma + \alpha I)$$

where Σ is the covariance of the data.

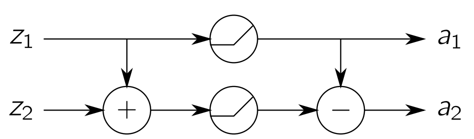
(b) *Show* that the condition number associated to this Hessian matrix is given by

$$c = \frac{\lambda_1 + \alpha}{\lambda_d + \alpha}$$

where $\lambda_1, \dots, \lambda_d$ are the eigenvalues of the matrix Σ sorted in decreasing order.

Exercise 3: Backpropagation (30 P)

Consider some portion of a neural network given by:



$$a_1 = \max(0, z_1)$$
$$a_2 = \max(0, z_1 + z_2) - a_1$$

(a) Assuming that we know the error gradient $(\partial E / \partial a_1 \, , \, \partial E / \partial a_2)$, *compute* the error gradient $(\partial E / \partial z_1 \, , \, \partial E / \partial z_2)$.