

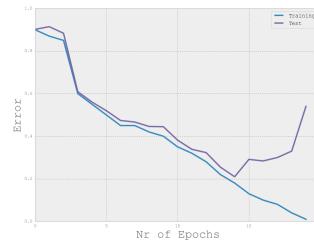
Deep Learning

Winter term 25/26 – Exercise Sheet 5

Submission Deadline: Monday, November 17, 2025, 2:00 PM

1. Early Stopping (2P)

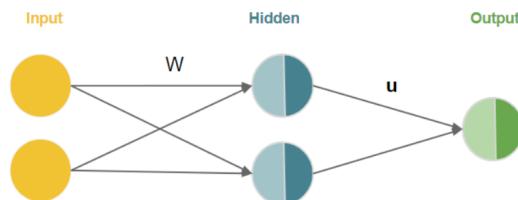
- a) Explain the early stopping strategy based on the following figure. (1P)



- b) Why does it make sense to look several steps ahead before deciding when to stop? (1P)

2. Dropout (8P)

Given is the following network with two input and two hidden neurons and ReLU activation functions in both hidden and output layer. The network parameters are given by the weight matrix $W = \begin{pmatrix} -1 & 2 \\ 3 & -4 \end{pmatrix}$, the weight vector $u = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, the bias of the hidden neuron $b = \begin{pmatrix} 3 \\ 3 \end{pmatrix}$, and the bias of the output neuron $c = 1$.



- a) Training: You want to train this neural network with dropout with a dropout probability of $p = 0.5$ and a mean-squared-error loss function. In the current mini-batch there are two training samples $x^{(1)} = (1, 1)$ and $x^{(2)} = (2, 2)$ with targets $y^{(1)} = 2$ and $y^{(2)} = 3$, respectively. In the first randomly generated dropout mask the first input neuron and the second hidden neuron are dropped out (i.e. the mask is given by $\mu^{(1)} = (0, 1, 1, 0)$ for our neurons (x_1, x_2, z_1, z_2)), and in the second mask only the first hidden neuron is dropped (i.e. $\mu^{(2)} = (1, 1, 0, 1)$). Perform one step of mini-batch stochastic gradient descent with learning rate $\alpha = 0.1$. (6P)
- b) Making a prediction: After training, given a new sample $x^{(3)} = (1, 2)$, calculate the networks prediction for that sample. (2P)