Homework 1: Deep Learning

Out April 18; Due May 1, 12 a.m.*

Kristian Kersting, Dominik Hintersdorf, Quentin Delfosse
{kersting, dominik.hintersdorf, quentin.delfosse}@cs.tu-darmstadt.de

1. Implement a feed-forward neural network architecture with the following constraints.

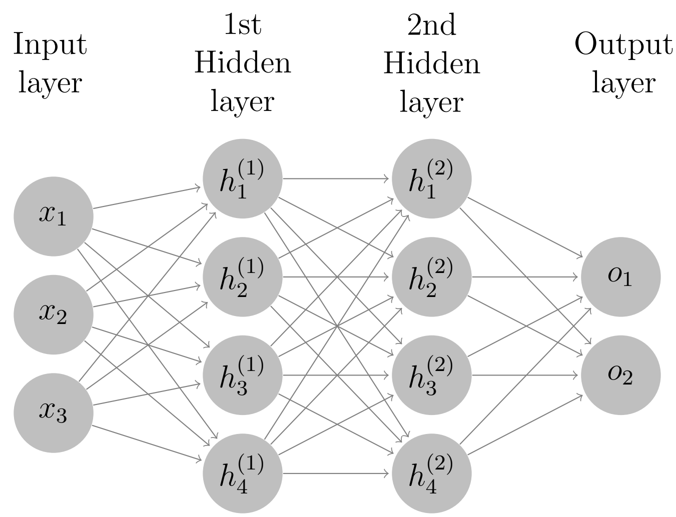

Figure 1: A feed-forward neural network architecture consisting of two hidden layers.

- The sigmoid activation function is used for computing hidden activations.
- No activation function in the output layer.
- In this homework, please use the attached notebook.

2. Given the forward pass of the above network architecture, please implement a mean squared loss function as follows

$$\mathcal{L}(\theta; \mathbf{x}, \mathbf{y}) = \frac{1}{M} \sum_{n=1}^{M} \frac{1}{2} \| f_\theta(\mathbf{x}_n) - \mathbf{y}_n \|^2 \tag{1}$$

where f_θ denotes the outputs of the network given inputs $\mathbf{x}$, $\mathbf{y}$ are the targets, $\mathbf{M}$ are the number of samples in the batch and θ denotes a set of the parameters.

*We will discuss the solutions in the exercise session. It is our suggestion that you try to address at least 50% of the exercise questions. Simply try hard to solve them. This way, you will get familiar with the technical terms and with the underlying ideas of the lecture.

3. Implement a *backward* pass of the network to propagate errors calculated in eq. (1).

4. Write a program that computes the numerical gradients for every parameter given the loss. For further details, see the lecture slides.

HINT: do not use back-propagation, just the forward evaluation implemented in the homework 1.

5. Please submit your notebook file to moodle that shows the following at the end:

- The numerical and backprop gradients of the loss function with respect to each parameter separately $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_1}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{b}_1}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_2}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{b}_2}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{W}_3}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{b}_3}$.
- A discussion on the relative error of the numerical and analytic gradients.