Neural Networks

Exercise 1 : Gradient Descent

(a) What are the differences between the perceptron training rule and the gradient descent method?

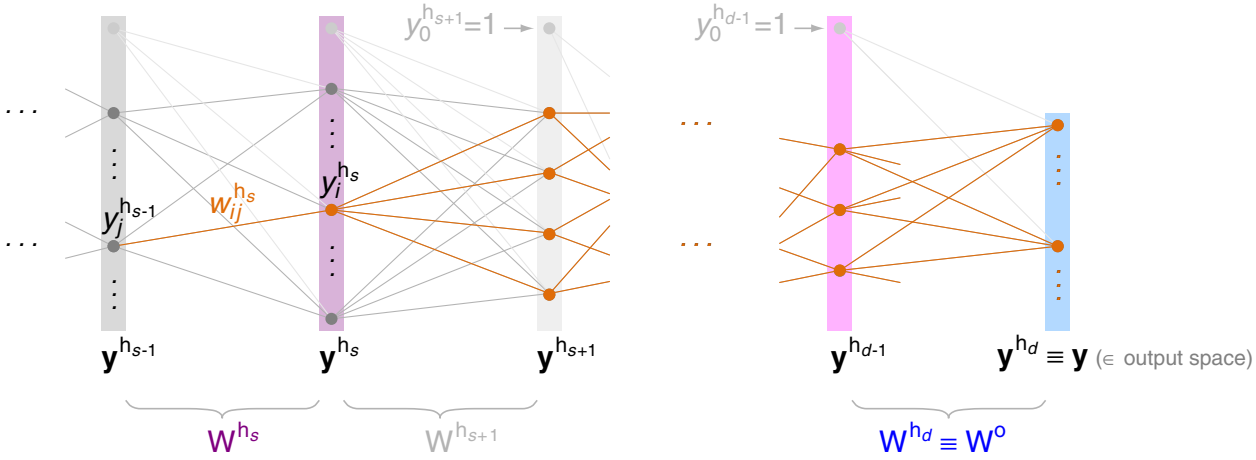
(b) What are the requirements for gradient descent being successful as a learning algorithm?

Exercise 2 : Perceptron Learning

How can perceptrons be applied to solve a classification problem with more than two classes?

Exercise 3 : From Chain Rule to Backpropagation

Consider the weight $w_{ij}^{h_s}$ in the following graph of a neural network:



In order to update the weight, we want to compute the derivative $\frac{\partial L(\mathbf{w})}{\partial w_{ij}^{h_s}}$.

- Verify the correctness of the following chain rule:

$$\frac{\partial L}{\partial w_{ij}^{h_s}} = \frac{\partial L}{\partial y_i^{h_s}} \cdot \frac{\partial y_i^{h_s}}{\partial v} \cdot \frac{\partial v}{\partial w_{ij}^{h_s}}$$

  with $v := \sum_j w_{ij}^{h_s} \cdot y_j^{h_{s-1}}$ and $y_i^{h_s} = \sigma(v)$.

- What is $\frac{\partial y_i^{h_s}}{\partial v}$?
- What is $\frac{\partial v}{\partial w_{ij}^{h_s}}$?
- $\frac{\partial L}{\partial y_i^{h_s}}$ can not be computed directly, but is based on results from a previous step. Identify terms from the network this value depends on.
- Which of those terms are computed in the forward propagation, which are computed in the backpropagation?