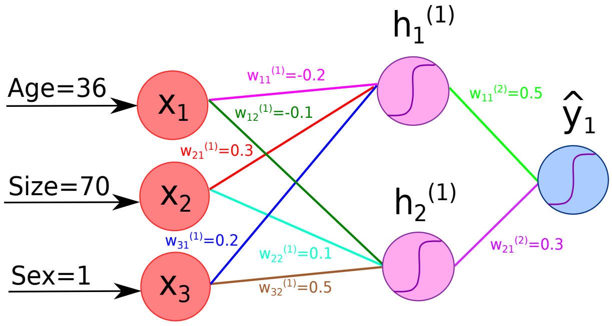
OTH Amberg-Weiden
Fakultät für Elektrotechnik, Medien und Informatik
Prof. Dr.-Ing. Christian Bergler, Prof. Dr. Fabian Brunner
Deep Learning – Winter Semester 2024/2025
Exercise for Lecture – Multi-Layer Perceptron

Task 1 (Multi-Layer Perceptron/Network)
Parameters, Shapes, and layers in a multi-layer Perceptron

a) Draw a multi-layer neural network consisting of three layers. The input layer consists of 4 nodes, the hidden layer of 5 nodes and the output layer of 3 nodes.

b) How many unknown parameters does the network have and what are the shapes of the individual vectors/matrices?

Task 2 (Multi-Layer Perceptron/Network)
The fully connected neural network (MLP) outlined below is used as a modelling approach for a regression task. For simplicity all bias values are equal to zero. If no other activation function is explicitly mentioned in the following, the sigmoid activation $\sigma(x) = \frac{1}{1+e^{-x}}$ is used for the entire MLP in all layers.



a) Lets assume the activation function across all layers is $a(x) = x$. Please show that the following equation $\hat{y} = f(x) = W^{(2)}W^{(1)}x$ applies if $a(x) = x$ is selected as the activation function! In addition, outline the relationship described by the network between the output $\hat{y}$ and the inputs $\boldsymbol{x} = (x_1, x_2, x_3)^T$ in vector form by adding the entries of the weight matrices $W^{(1)}$ and $W^{(2)}$ to the following representation:

$$\hat{y} = f(\boldsymbol{x}) = W^{(2)}W^{(1)}\boldsymbol{x} = \Big(\qquad\qquad \Big) \Bigg(\qquad\qquad\qquad \Bigg) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

b) Describe the network dimensionality (parametric complexity) and how they can be reduced, by drawing an equivalent model with as few nodes as possible, which can represent the same functionality as $\hat{y} = f(x) = W^{(2)}W^{(1)}x$.

Now let us also have an annotated training data set of length m consisting of samples $(\boldsymbol{x}^{(i)}, y^{(i)})$, $i \in \{1, \ldots, m\}$, $\boldsymbol{x}^{(i)} \in \mathbb{R}^3, y^{(i)} \in \mathbb{R}$. To determine the weights, the loss function

$$L(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}^{(i)} - y^{(i)})^2$$

where $\hat{y}^{(i)} = f(\boldsymbol{x}^{(i)})$ denotes the model output for the input $\boldsymbol{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, x_3^{(i)})^T$.

c) Explain the task and the principle of the backpropagation algorithm in model training.

d) Compute the forward propagation (vectorized form) for the above visualized MLP, using the given input feature vector $x = (36, 70, 1)^T$ with $m = 1$ (single sample) and sigmoid activations $\sigma(x)$, together with the mentioned loss function $L(\theta)$ and a ground truth value $y = 1$.

e) Compute the backward propagation (vectorized form) for the given MLP, in order to derive all the gradients required for a complete parameter update.