Introduction to Machine Learning

Chapter 3: Deep Learning- MLP without LA

Bernd Bischl

Department of Statistics – LMU Munich

Winter term 2021

Multi-Layer Feedforward Neural Networks

FEEDFORWARD NEURAL NETWORKS

- We will now extend the model class once again, such that we allow an arbitrary amount of l (hidden) layers.
- The general term for this model class is (multi-layer) **feedforward networks** (inputs are passed through the network from left to right, no feedback-loops are allowed)

FEEDFORWARD NEURAL NETWORKS

- We can characterize those models by the following chain structure:

  $$f(\mathbf{x}) = \tau \circ \phi \circ \sigma^{(l)} \circ \phi^{(l)} \circ \sigma^{(l-1)} \circ \phi^{(l-1)} \circ \ldots \circ \sigma^{(1)} \circ \phi^{(1)}$$

  where $\sigma^{(i)}$ and $\phi^{(i)}$ are the activation function and the weighted sum of hidden layer i, respectively. τ and ϕ are the corresponding components of the output layer.

- Each hidden layer has:
  - an associated weight matrix $\mathbf{W}^{(i)}$, bias $\mathbf{b}^{(i)}$ and activations $\mathbf{z}^{(i)}$ for $i \in \{1 \ldots l\}$
  - $\mathbf{z}^{(i)} = \sigma^{(i)}(\phi^{(i)}) = \sigma^{(i)}(\mathbf{W}^{(i)T}\mathbf{z}^{(i-1)} + \mathbf{b}^{(i)})$, where $\mathbf{z}^{(0)} = \mathbf{x}$.

- Again, without non-linear activations in the hidden layers, the network can only learn linear decision boundaries.

FEEDFORWARD NEURAL NETWORKS

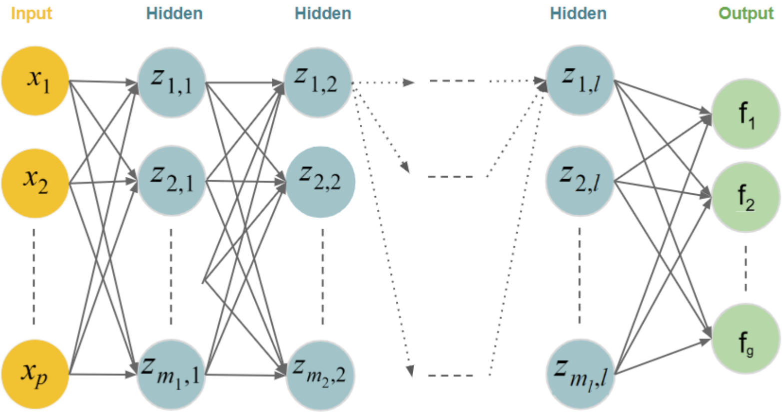
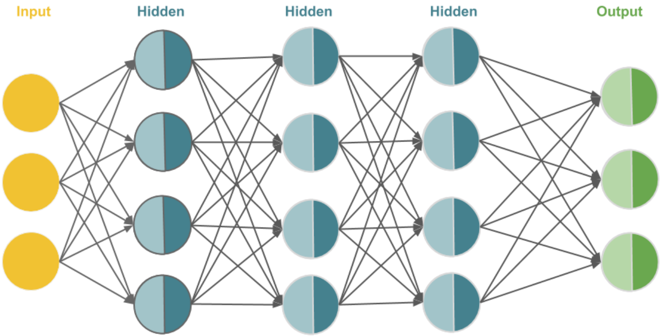


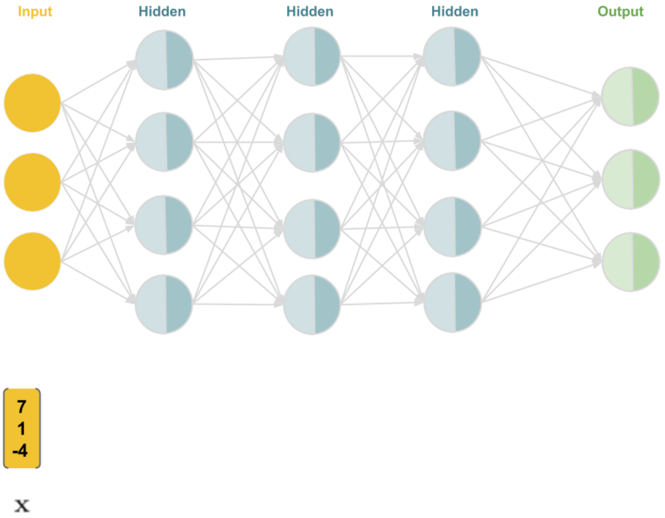
Figure: Structure of a deep neural network with *l* hidden layers (bias terms omitted).

FEEDFORWARD NEURAL NETWORKS: EXAMPLE

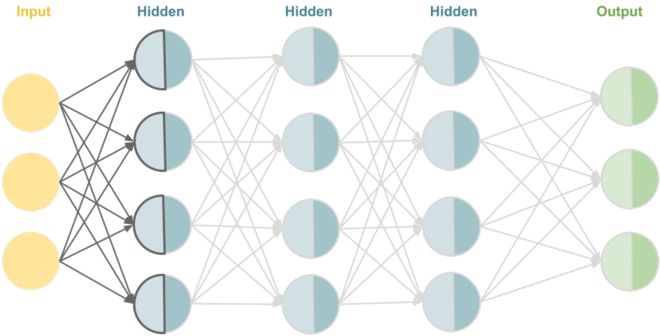


$$\underbrace{\begin{pmatrix} 13 & -9 & 2 \\ -8 & 0 & 3 \\ 4 & -1 & 5 \\ -3 & 12 & 7 \end{pmatrix}}_{(W^{(1)})^T} \underbrace{\begin{pmatrix} 5 \\ -2 \\ 2 \\ 11 \end{pmatrix}}_{\mathbf{b}^{(1)}} \underbrace{\begin{pmatrix} 1 & 0 & -4 & 1 \\ 0 & 11 & 2 & -14 \\ -1 & 5 & -2 & 16 \\ 0 & -9 & -3 & 4 \end{pmatrix}}_{(W^{(2)})^T} \underbrace{\begin{pmatrix} -5 \\ 3 \\ 1 \\ -8 \end{pmatrix}}_{\mathbf{b}^{(2)}} \underbrace{\begin{pmatrix} 1 & -2 & -18 & -7 \\ 3 & -4 & 8 & 0 \\ -2 & 1 & 21 & 5 \\ 2 & -2 & 11 & -13 \end{pmatrix}}_{(W^{(3)})^T} \underbrace{\begin{pmatrix} 4 \\ -6 \\ 1 \\ -17 \end{pmatrix}}_{\mathbf{b}^{(3)}} \underbrace{\begin{pmatrix} 9 & 3 & -1 & -4 \\ -8 & -2 & 14 & 3 \\ 13 & 2 & -9 & -1 \end{pmatrix}}_{U^T} \underbrace{\begin{pmatrix} -1 \\ -4 \\ -30 \end{pmatrix}}_{\mathbf{c}}$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE

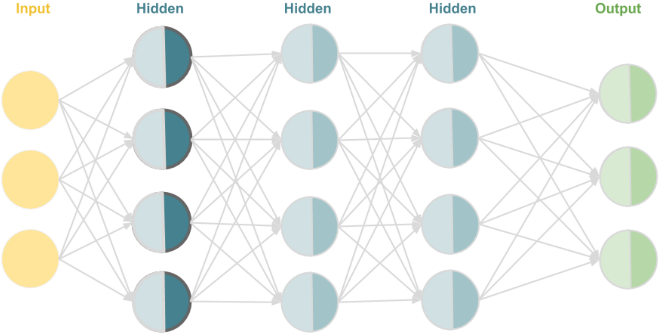


FEEDFORWARD NEURAL NETWORKS: EXAMPLE



$$\underbrace{\begin{pmatrix} 7 \\ 1 \\ -4 \end{pmatrix}}_{\mathbf{x}} \quad \underbrace{\begin{pmatrix} 7*13 + 1*(-9) + (-4)*2 + 5 \\ 7*(-8) + 1*0 + (-4)*3 + (-2) \\ 7*4 + 1*(-1) + (-4)*5 + 2 \\ 7*(-3) + 1*12 + (-4)*7 + 11 \end{pmatrix}}_{\mathbf{z}_{in}^{(1)} = W^{(1)T}\mathbf{x} + \mathbf{b}^{(1)}}$$

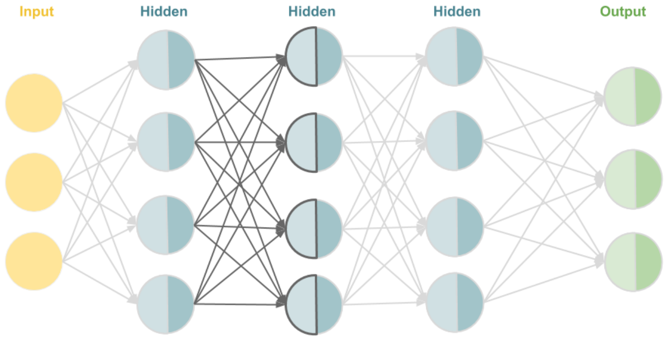
FEEDFORWARD NEURAL NETWORKS: EXAMPLE



$$\begin{pmatrix} 7 \\ 1 \\ -4 \end{pmatrix} \quad \begin{pmatrix} 79 \\ -70 \\ 9 \\ -26 \end{pmatrix} \quad \begin{pmatrix} \max(0, 79) \\ \max(0, -70) \\ \max(0, 9) \\ \max(0, -26) \end{pmatrix}$$

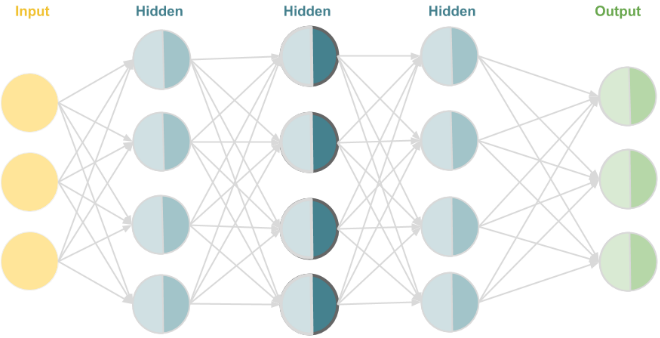
$$\mathbf{x} \qquad \mathbf{z}_{in}^{(1)} \quad \mathbf{z}^{(1)} = \mathbf{z}_{out}^{(1)} = \sigma(\mathbf{z}_{in}^{(1)})$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE



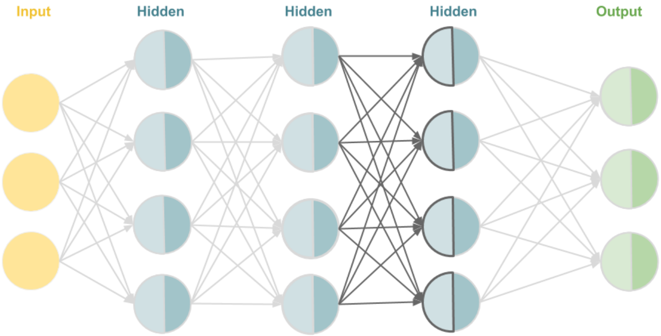
$$\underbrace{\begin{pmatrix} 7 \\ 1 \\ -4 \end{pmatrix}}_{\mathbf{x}} \quad \underbrace{\begin{pmatrix} 79 \\ -70 \\ 9 \\ -26 \end{pmatrix}}_{\mathbf{z}_{in}^{(1)}} \quad \underbrace{\begin{pmatrix} 79 \\ 0 \\ 9 \\ 0 \end{pmatrix}}_{\mathbf{z}^{(1)}} \quad \underbrace{\begin{pmatrix} 79*1 + 0*0 + 9*(-4) + 0*1 + (-5) \\ 79*0 + 0*11 + 9*2 + 0*(-14) + 3 \\ 79*(-1) + 0*5 + 9*(-2) + 0*16 + 1 \\ 79*0 + 0*(-9) + 9*(-3) + 0*4 + (-8) \end{pmatrix}}_{\mathbf{z}_{in}^{(2)} = W^{(2)T}\mathbf{z}^{(1)} + \mathbf{b}^{(2)}}$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE



$$\mathbf{x} = \begin{bmatrix} 7 \\ 1 \\ -4 \end{bmatrix} \quad \mathbf{z}_{in}^{(1)} = \begin{bmatrix} 79 \\ -70 \\ 9 \\ -26 \end{bmatrix} \quad \mathbf{z}^{(1)} = \begin{bmatrix} 79 \\ 0 \\ 9 \\ 0 \end{bmatrix} \quad \mathbf{z}_{in}^{(2)} = \begin{bmatrix} 38 \\ 21 \\ -96 \\ -35 \end{bmatrix} \quad \mathbf{z}^{(2)} = \mathbf{z}_{out}^{(2)} = \sigma(\mathbf{z}_{in}^{(2)}) = \begin{pmatrix} \max(0, 38) \\ \max(0, 21) \\ \max(0, -96) \\ \max(0, -36) \end{pmatrix}$$

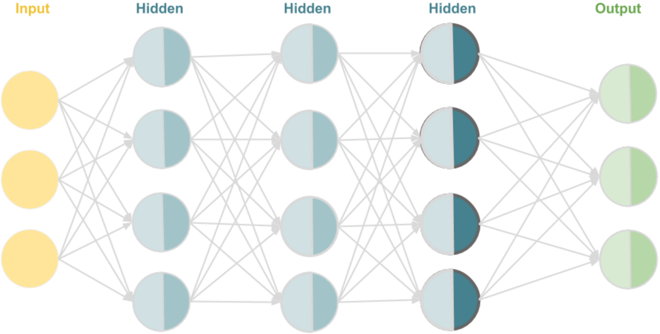
FEEDFORWARD NEURAL NETWORKS: EXAMPLE



$$\mathbf{x} = \begin{bmatrix}7\\1\\-4\end{bmatrix} \quad \mathbf{z}_{in}^{(1)} = \begin{bmatrix}79\\-70\\9\\-26\end{bmatrix} \quad \mathbf{z}^{(1)} = \begin{bmatrix}79\\0\\9\\0\end{bmatrix} \quad \mathbf{z}_{in}^{(2)} = \begin{bmatrix}38\\21\\-96\\-35\end{bmatrix} \quad \mathbf{z}^{(2)} = \begin{bmatrix}38\\21\\0\\0\end{bmatrix}$$

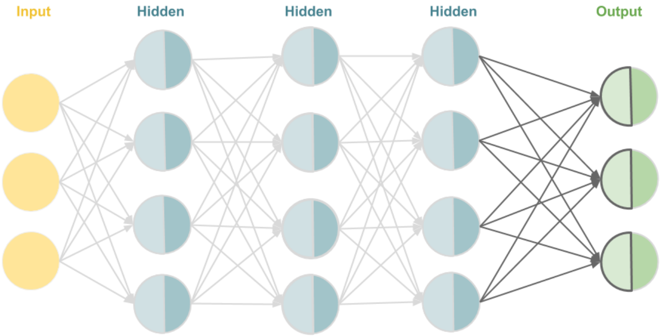
$$\mathbf{z}_{in}^{(3)} = W^{(3)T}\mathbf{z}^{(2)} + \mathbf{b}^{(3)} = \begin{pmatrix} 38*1 + 21*(-2) + 0*(-18) + 0*(-7) + 4 \\ 38*3 + 21*(-4) + 0*8 + 0*0 + (-6) \\ 38*(-2) + 21*1 + 0*21 + 0*5 + 1 \\ 38*2 + 21*(-2) + 0*11 + 0*(-13) + (-17) \end{pmatrix}$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE



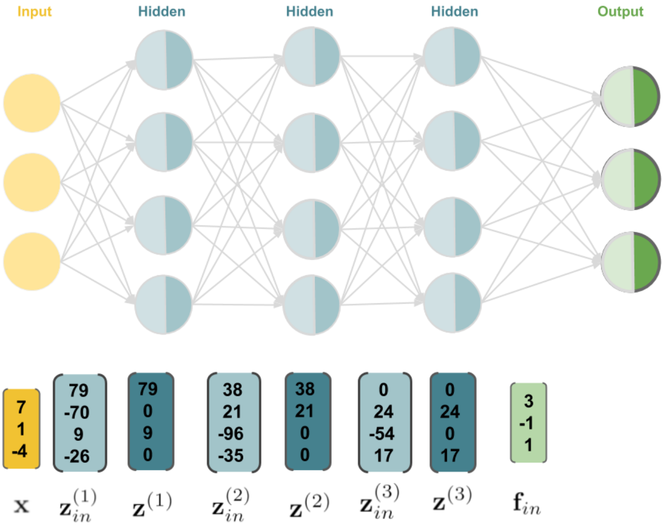
$$
\begin{array}{ccccccc}
\begin{pmatrix}7\\1\\-4\end{pmatrix} &
\begin{pmatrix}79\\-70\\9\\-26\end{pmatrix} &
\begin{pmatrix}79\\0\\9\\0\end{pmatrix} &
\begin{pmatrix}38\\21\\-96\\-35\end{pmatrix} &
\begin{pmatrix}38\\21\\0\\0\end{pmatrix} &
\begin{pmatrix}0\\24\\-54\\17\end{pmatrix} &
\begin{pmatrix}\max(0, 0)\\\max(0, 24)\\\max(0, -54)\\\max(0, 17)\end{pmatrix} \\
\mathbf{x} & \mathbf{z}_{in}^{(1)} & \mathbf{z}^{(1)} & \mathbf{z}_{in}^{(2)} & \mathbf{z}^{(2)} & \mathbf{z}_{in}^{(3)} & \mathbf{z}^{(3)} = \mathbf{z}_{out}^{(3)} = \sigma(\mathbf{z}_{in}^{(3)})
\end{array}
$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE

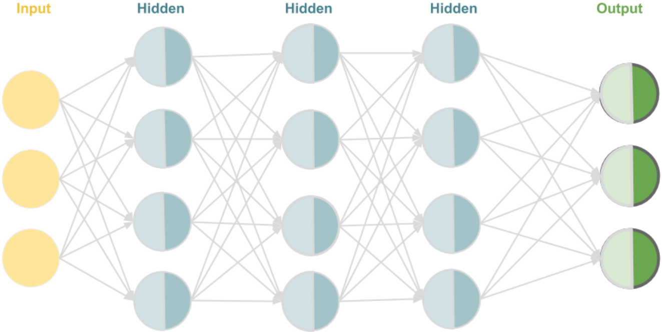


$$\underset{\mathbf{x}}{\begin{pmatrix} 7 \\ 1 \\ -4 \end{pmatrix}} \quad \underset{\mathbf{z}_{in}^{(1)}}{\begin{pmatrix} 79 \\ -70 \\ 9 \\ -26 \end{pmatrix}} \quad \underset{\mathbf{z}^{(1)}}{\begin{pmatrix} 79 \\ 0 \\ 9 \\ 0 \end{pmatrix}} \quad \underset{\mathbf{z}_{in}^{(2)}}{\begin{pmatrix} 38 \\ 21 \\ -96 \\ -35 \end{pmatrix}} \quad \underset{\mathbf{z}^{(2)}}{\begin{pmatrix} 38 \\ 21 \\ 0 \\ 0 \end{pmatrix}} \quad \underset{\mathbf{z}_{in}^{(3)}}{\begin{pmatrix} 0 \\ 24 \\ -54 \\ 17 \end{pmatrix}} \quad \underset{\mathbf{z}^{(3)}}{\begin{pmatrix} 0 \\ 24 \\ 0 \\ 17 \end{pmatrix}} \quad \underset{\mathbf{f}_{in} = U^T \mathbf{z}^{(3)} + \mathbf{c}}{\begin{pmatrix} 24 \cdot 3 + 17 \cdot (-4) + (-1) \\ 24 \cdot (-2) + 17 \cdot 3 + (-4) \\ 24 \cdot 2 + 17 \cdot (-1) + (-30) \end{pmatrix}}$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE



FEEDFORWARD NEURAL NETWORKS: EXAMPLE



$$\mathbf{f}_{in} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

$$\mathbf{f} = \mathbf{f}_{out} = \tau(\mathbf{f}_{in}) = \begin{pmatrix} \frac{\exp(3)}{\exp 3 + \exp(-1) + \exp(1)} \\ \frac{\exp(-1)}{\exp 3 + \exp(-1) + \exp(1)} \\ \frac{\exp(1)}{\exp 3 + \exp(-1) + \exp(1)} \end{pmatrix}$$

FEEDFORWARD NEURAL NETWORKS: EXAMPLE