# Neural Network Theory (3+1 hrs/week, 5 ECTS)

Cornelia Schneider cornelia.schneider@fau.de

Summer term 2026

**Contents:**

- **Classical approximation results by neural networks** (universality, approximation of smooth functions)
- **ReLU networks** (linear finite elements, approximation of the square function)
- **The role of depth** (representation of compactly supported functions, number of pieces)
- **High dimensional approximation** (curse of dimensionality, dimension dependent regularity assumptions)
- **Complexity of sets of networks** (VC dimension, lower bounds)
- **Spaces of realisations of neural networks** (non-convexity and closedness of network spaces)

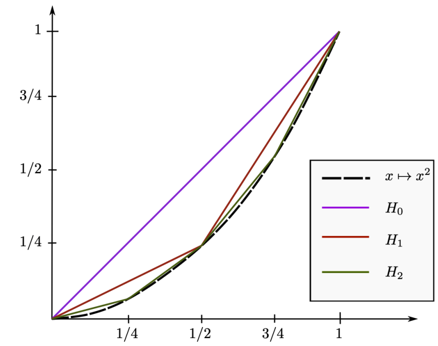


Fig. 1: Approximation of $x \mapsto x^2$.

**Description:** In this lecture, we will explore the mathematical structure of neural networks (NNs), a central concept in modern machine learning. Historically, the first model was inspired by biology: the McCulloch and Pitts neuron (1943) is a simple function which mimics a biological neuron firing when its weighted inputs exceed a threshold.

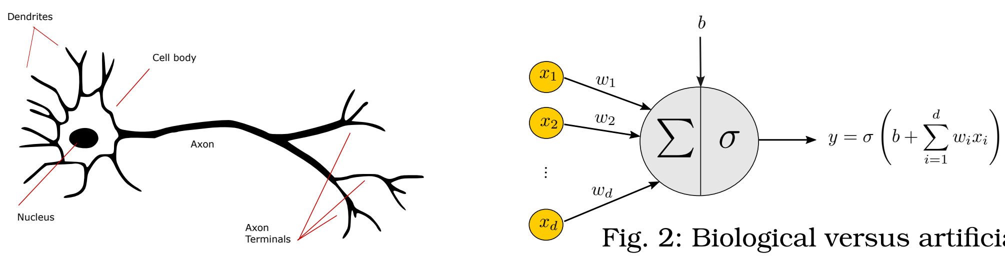


Fig. 2: Biological versus artificial neuron.

A network of neurons can be constructed by linking multiple neurons together such that the output of one neuron forms an input to another. A simple model for such a network is the multilayer perceptron (MLP). An MLP with $L$ layers and activation $\sigma$ is a function of the form

$$F(x) = T_L(\sigma(T_{L-1}(\ldots\sigma(T_1(x))\ldots))),$$

where each $T_l x := A_l x + b_l$, $l = 1, \ldots, L$ is an affine map and $\sigma : \mathbb{R} \to \mathbb{R}$ a (nonlinear) activation function acting coordinate-wise. Although this architecture – with connections only between adjacent layers – is far from a faithful biological model, it forms the computational core of deep learning.

The focus of this lecture will be mathematical, not algorithmic. We will study MLPs from a functional-analytic perspective. By concentrating on a rigorous analytical framework, we will uncover mathematical reasons behind many empirically observed successes of deep learning, providing a foundation to understand their strengths and limitations. In particular, we address questions such as:

- Why are deep networks often more powerful than shallow ones?
- How can NNs approximate high-dimensional functions efficiently, where classical methods fail?
- What structural properties make certain architectures preferable?

**Prerequisites:** Linear Algebra I and Analysis I-III or Mathematics for Data Science I+II or comparable. Moreover, Functional Analysis or Higher Analysis are recommended. This course is suitable for Bachelor students in Mathematics as well as for the Master Mathematics/Data Science/CAM.

**Creditable as:** Theoretical/Applied Mathematics (B.Sc. Math); Specialisation Mathematical Theory/Foundations of Data Science (Data Science); Specialization: Analysis and Stochastics (M.Sc. Math).

## Literature

[MP16] H.N. Mhaskar and T. Poggio. Deep vs. shallow networks: An approximation theory perspective. *Analysis and Applications*, 14(06):829–848, 2016.

[ZP24] P. Petersen and J. Zech. Mathematical theory of deep learning. *arXiv: 2407.18384v2.*

[Rud91] W. Rudin. *Functional Analysis* McGraw-Hill, 1991.

[Yar17] D. Yarotsky. Error bounds for approximations with deep ReLU networks. *Neural Networks*, 94:103–114, 2017.