Notes on Deep Learning: Mathematical Exposition

Compiled by D. Gueorguiev, 4/30/2025

Deep learning contains the following sub-disciplines:

Deep Supervised learning

Deep Unsupervised learning

Deep Reinforcement learning

We will consider the first discipline – Deep Supervised Learning

# Deep Supervised Learning: Problem Formulation

Let , ,

Let. (pf.1)

is the number of available input-output data pairs. is the dimension of the input data.

is an unknown function which relates input and output data through (pf.1).

is the available input data. is the available output data.

**Definition**: *Learning problem of type*

The objective is to approximately compute the output of the -th input data point without explicit knowledge of the function but instead using the knowledge of the input-output data pairs

, (pf.2)

To accomplish this we consider the optimization problem of computing approximate minimizers of the function defined for such that

(pf.3)

Clearly, (pf.1) ensures that and the unknown function is a minimizer of .

Note that is defined on the infinite dimensional vector space , given the fact that functional spaces in general can be viewed as infinite dimensional vector spaces. Therefore, we introduce a spatially discretized version of this optimization problem:

Let , be a function, and let satisify

(pf.5)

We think of the set

(pf.6)

as a parametrized set of functions which we employ to approximate the infinite dimensional vector space .

Thus we can think of the function

(pf.7)

as the parametrization function associated with this set .

For example, in the case one could think of (pf.7) as the sense that for all it holds that

(pf.8)

Also one can think of (pf.7) as the parametrization associated to trigonometric polynomials (i.e. Fourier series expansion).

In Deep Supervised Learning one chooses a parametrization associated to deep ANNs rather than parametrizations on polynomials or trigonometric polynomials.

Taking the set in (pf.6) and its parametrization function in (pf.7) into account, we then want to compute approximate minimizers of the function restricted to the set , that is we consider the optimization problem of computing approximate minimizers of the function

(pf.9)

Employing the parametrization function (pf.7) one reformulates the optimization problem (pf.9) as an optimization problem of computing approximate minimizers of the function

(pf.10)

(pf.10) is suitable for implementing numerical algorithms for solving it. In the context of Deep Supervised Learning, one chooses the parametrization function in (pf.7) as deep ANN parametrizations to which one would apply SGD-type optimization algorithm for computing the approximate minimizers of (pf.10).

We will discuss the most common variants of such SGD-type optimization algorithms. If is an approximate minimizer of (pf.10) in the sense that , one then considers as an approximation

(pf.11)

of the unknown output of the -th input data point .

We note that in Deep Supervised Learning algorithms one typically aims to compute an approximate minimizer of (pf.10) in the sense that which typically is not a minimizer of (pf.10).

In (pf.3) above we have set up an optimization problem for the learning problem by using the standard mean squared error function to measure the loss. This *mean squared error loss function* is just one possible example in the formulation of deep learning optimization problems. In particular, in image classification problems and NLP problems other loss functions such as *the cross-entropy loss function* are often used.

# Artificial Neural Nets (ANNs)

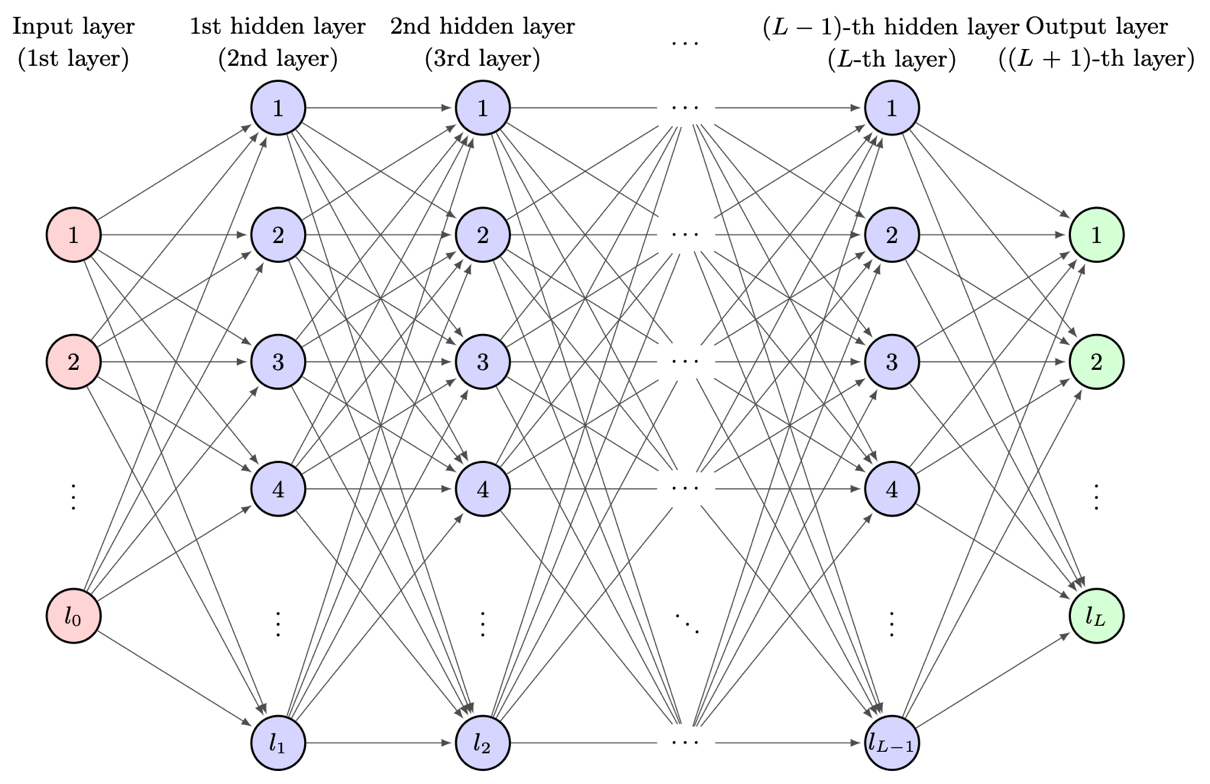


Figure 1: Graphical illustration of a fully-connected feedforward ANN consisting of affine transformations (i.e. consisting of layers: one input layer, hidden layers, and one output layer) with neurons on the input layer (i.e. with -dimensional input layer), with neurons on the -th hidden layer, and neurons in the output layer.

## Affine Functions

**Definition**: *Affine functions*

Let satisfy . Then we denote by the function which satisfies for all that

(ann.1)

is the affine function associated with

## Vectorized description of fully-connected feedforward networks

**Definition**: *Vectorized description of fully connected feedforward network*

Let satisfy

(ann.2)

Let for be a function. Then we denote by the function which satisfies for all that

# Feedforward Neural Networks

//TODO: lay out exposition based on Ch 2 of [2]

# References

[1] [Mathematical Introduction to Deep Learning: Methods, Implementations, and Theory, Arnulf Jentzen, Benno Kuckuck, Phillipe von Wurstemberger, University of Muenster, 2023](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/books/Mathematical_Introduction_to_Deep_Learning-Methods_Implementations_and_Theory_Jentzen_2023.pdf)

[2] [Mathematical Theory of Deep Learning, Philipp Petersen, Jakob Zech, 2024](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/books/Mathematical_theory_of_deep_learning_Petersen_2024.pdf)

[3] [The Theory of Deep Learning, Sanjeev Arora, Princeton U., 2022](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/books/Theory_of_Deep_Learning_Sanjeev_Arora.pdf)

[4] [The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks, Daniel A. Roberts, Sho Yaida, Boris Hanin, 2021](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/books/The_Principles_of_Deep_Learning_Theory_Roberts_2021.pdf)

[5] [Foundations of Machine Learning, Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, 2012](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/books/Foundations_of_Machine_Learning_Mohri_2012.pdf)