Notes on the Theory of Networks for Approximation and Learning

compiled by D.Gueorguiev, 4/11/2025

# Introductory Notes

Learning an input-output mapping from a set of examples, of the type that many neural nets have been constructed to perform, can be regarded as synthesizing an approximation of a multi-dimensional function, that is solving the problem of hypersurface reconstruction. From this point of view this form of learning is closely related to classical approximation techniques, such as generalized splines and regularization theory. We would like to investigate

# References

[1] [A Theory of Networks for Approximation and Learning, T. Poggio et al, 1989](https://github.com/dimitarpg13/statistical_learning_and_kernel_methods/blob/main/literature/articles/learning_as_approximation/A_Theory_of_Networks_for_Approximation_and_Learning_Poggio_Girosi_1989.pdf)

[2] [The Statistical Mechanics of Learning A Rule, T. Watkin et al, 1993](https://github.com/dimitarpg13/statistical_learning_and_kernel_methods/blob/main/literature/articles/learning_as_approximation/The_Statistical_Mechanics_of_Learning_A_Rule_1993RevModPhysWatkin.pdf)

[3] [Statistical Mechanics of Learning, Andreas Engel, 1999](https://github.com/dimitarpg13/statistical_learning_and_kernel_methods/blob/main/literature/articles/learning_as_approximation/Statistical_Mechanics_of_Learning_Engel_1999.pdf)

[4] [The Peaking Phenomenon in Semi-Supervised Learning, JH Krijte, M. Loog, 2016](https://github.com/dimitarpg13/statistical_learning_and_kernel_methods/blob/main/literature/articles/learning_as_approximation/The_Peaking_Phenomenon_in_Semi-supervised_Learning_Krijthe_2016.pdf)

[5] [Reconciling modern machine learning practice and the bias-variance trade-off, Mikhail Belkin et al, 2019](https://github.com/dimitarpg13/statistical_learning_and_kernel_methods/blob/main/literature/articles/learning_as_approximation/Reconciling_modern_machine_learning_practice_and_the_bias-variance_trade-off_Belkin_2019.pdf)

[6] [Kolmogorov-Arnold representation theorem, Wikipedia](https://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Arnold_representation_theorem)

# Appendix

## Kolmogorov-Arnold Representation Theorem

**Theorem** : *Kolmogorov-Arnold function representation*

Every multivariate continuous function can be represented as a superposition of continuous single-variable functions. Precisely, if is a multivariate continuous function, then can be written as a finite composition of continuous functions of a single variable and the binary operation of addition.

where and

George Lorentz provides another formulation of this theorem in 1962 which shows that the outer functions can be replaced by a single function . More precisely,

David Sprecher replaced the inner functions by one single inner function with an appropriate shift in its argument. He proved that there exist real values , a continuous function , and a real increasing continuous function with

//TODO: finish the appendix section on Kolmogorov representation theorem

## Lipschitz-continuity of functions

Given two metric spaces and , where denotes the metric on the set and denotes the metric on the set , a function is called Lipschitz continuous if there exist a real constant such that, for and in ,

Any such K is referred to as a Lipschitz constant for the function and may also be referred as -Lipschitz (also denoted as .