An Introduction to Deep Learning Theory (M12)

Non-Examinable (Graduate Level)

Tianyi Ma

Deep learning has become one of the central tools of modern artificial intelligence, with remarkable empirical success in areas such as computer vision, natural language processing and generative modelling. However, many aspects of its behaviour remain mathematically subtle: why neural networks often outperform classical methods in high-dimensional tasks, why they generalise beyond the training data, and why large neural networks can be trained effectively despite the non-convex nature of the underlying optimisation problem. This course provides a brief introduction to the theoretical aspects of deep learning, focusing on approximation, optimisation, and statistical properties of neural networks.

The course will cover a selection of topics including:

- Statistical learning theory and approximation theory for neural networks.
- Training dynamics of neural networks: Neural Tangent Kernel (NTK) and mean-field dynamics.
- An overview of large language models and transformers.

Prerequisites

Basic knowledge of statistics, probability, linear algebra and analysis.

Literature

1. Schmidt-Hieber, J. (2020). Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics*
2. Bach, F. (2024). Learning theory from first principles. *MIT press*. (Chapters 9 and 12 are very useful)
3. Misiakiewicz, T. and Montanari, A. (2023) Six lectures on linearized neural networks. *arXiv preprint arXiv:2308.13431*.
4. Hannin, B. (2026) Neural Networks: a Primer for Mathematicians. Available at https://hanin.princeton.edu/nn-notes.pdf.
5. Fan, J., Ma, C., Zhong, Y. (2021) A Selective Overview of Deep Learning. *Statistical Science*.

Additional support

Printed notes will be available on Moodle.