Topics in Mathematics for Deep Learning (L16)

Non-Examinable (Graduate Level)

Dr C. Esteve Yagüe

This non-examinable course is intended to be an introduction to the mathematical theory of Deep Learning. The course is aimed at students with a mathematical background, with the goal of attracting their attention to this new research field which is nowadays experiencing a huge development. The course can also be interesting for computer scientists, with experience (or not) in deep learning, who want to explore the topic from a mathematical viewpoint.

The great success of the application of deep learning techniques to a wide range of real-life problems has raised a number of mathematical questions, some of them being still unanswered, or partially answered. We will start by introducing the notation used to describe artificial neural networks and formulate the underlying learning problems. We will review the most important theoretical results concerning the approximation and generalization properties of deep neural networks in the overparametrized regime (infinite width and depth limit), as well as the behaviour of the optimization algorithms used during training (typically formulated as a non-convex optimization problem). Then, we will discuss the use of different NN architectures, which guarantee desirable properties of the learned model. Finally, we will present some of the applications of deep learning techniques, such as inverse problems, image analysis, and the numerical approximation of partial differential equations.

The course will cover a selection of topics including:

- **Introduction to Deep Learning:** formulation of the learning problem, neural network architectures and types of error.
- **Expressivity of Neural Networks:** universal approximation of overparametrized neural networks in different functional spaces.
- **Optimization of Deep Neural Networks:** convergence of stochastic gradient descent, loss landscape analysis, implicit regularization, etc.
- **Generalization of Neural Networks:** connection to kernel methods, Gaussian processes and the kernel regime.
- **Special architectures:** residual NNs, convolutional NNs, scattering transform, equivariant NNs, etc.
- **Applications of deep learning:** Physics Informed NNs, Deep Learning for inverse problems, etc.

Prerequisites

No previous knowledge on deep learning or machine learning is required for this course, which is mainly addressed to students with a background in mathematics. Concerning the mathematical knowledge, it is desirable to have some fundamental background on functional analysis, measure theory and optimization.

Preliminary Reading

1. Julius Berner, Philipp Gtohs, Gitta Kutyniok, Philipp Petersen, *The Modern Mathematics of Deep Learning*, 2022. Book chapter in *Theory of Deep Learning*, Cambridge University Press. https://arxiv.org/abs/2105.04026

Literature

1. Francis Bach, *Learning theory from first principles*, 2021. Online version https://www.di.ens.fr/~fbach/ltfp_book.pdf

2. F. Bach, *Breaking the curse of dimensionality with convex neural networks*, The Journal of Machine Learning Research, 18(1):629–681, 2017. Link to journal https://jmlr.org/papers/volume18/14-546/14-546.pdf

3. J. E. Gerken, J. Aronsson, O. Carlsson, H. Linander, F. Ohlsson, C. Petersson, and D. Persson, *Geometric deep learning and equivariant neural networks*, Artificial Intelligence Review, pages 1–58, 2023. Link to journal https://link.springer.com/content/pdf/10.1007/s10462-023-10502-7.pdf

4. A. Jacot, F. Gabriel, and C. Hongler, *Neural tangent kernel: Convergence and generalization in neural networks*, Advances in neural information processing systems, 31, 2018. Link to journal https://proceedings.neurips.cc/paper_files/paper/2018/file/5a4be1fa34e62bb8a6ec6b91d2462f5a-Paper.pdf

5. A. G. de G. Matthews, J. Hron, M. Rowland, R. E. Turner, and Z. Ghahramani, *Gaussian process behaviour in wide deep neural networks*, In International Conference on Learning Representations, 2018 Link to journal https://arxiv.org/abs/1804.11271

6. S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb. *Learned reconstruction methods with convergence guarantees: A survey of concepts and applicattions.* IEEE Signal Processing Magazine, 40(1):164–182, 2023. Link to journal https://ieeexplore.ieee.org/abstract/document/10004773

Additional support

Lecture notes will be provided and accompanying office hours will be offered.