Seminar: Approximation Theory for Neural Networks

Prof. Jakob Zech

jakob.zech@uni-heidelberg.de

April 12, 2020

This seminar discusses the mathematical theory of function approximation using neural networks. We review classical results such as the universal approximation theorem, and more recent advances concerning provably attainable and optimal convergence rates. The convergence rates typically are valid for all elements in a class of functions over a (compact) domain $D \subset \mathbb{R}^d$, and depend on the regularity of the function class as well as the dimension d. Particular focus will be on the ReLU activation function and high-dimensional domains.

Each student will present one publication and write a report summarizing the key results and proofs of the paper. All presentations and meetings take place online, more details will be published on the Moodle page before the first meeting.

First meeting: Wednesday, April 22, 11:15 to 13:00

Tentative list of topics and references (the $*$ indicates that this is supplementary material not directly related to neural networks):

- **Universal Approximation Theorems**
  - G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signals Systems*, 1989. (also see the erratum for a correction of the proof)
  - K. Hornik, Approximation Capabilities of Multilayer Feedforward Networks, *Neural Networks*, 1991.
  - H. N. Mhaskar and C. A. Micchelli, Approximation by superposition of sigmoidal and radial basis functions, *Advances in Applied Mathematics*, 1992.
- **Approximation of polynomials and rational functions**
  - D. Yarotksy, Error bounds for approximations with deep ReLU networks , 2016.
  - M. Telgarsky, Neural networks and rational functions, 2017.
- **Piecewise linear approximation**
  - R. Arora, A. Basu, P. Mianjy, A. Mukherjee, Understanding Deep Neural Networks with Rectified Linear Units, 2016.

- J. He, L. Li, J. Xu, C. Zheng, ReLU Deep Neural Networks and Linear Finite Elements, 2018.

- **Approximation in L^p**
  - P. Petersen, F. Voigtlaender, Optimal approximation of piecewise smooth functions using deep ReLU neural networks, 2017.

- **Curse of dimensionality**
  - A.R. Barron, Universal approximation bounds for superpositions of a sigmoidal function, *IEEE Transactions on Information Theory*, 1993.
  * E. Novak, H. Wozniakowski, Approximation of infinitely differentiable multivariate functions is intractable, *Journal of Complexity*, 2009.
  - C. Schwab and J. Zech, Deep learning in high dimension: Neural network expression rates for generalized polynomial chaos expansions in UQ., *Analysis and Applications*, 2017.
  - T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, Q. Liao, Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review, *International Journal of Automation and Computing*, 2017.

- **Benefits of depth**
  - R. Eldan, O. Shamir, The Power of Depth for Feedforward Neural Networks, *JMLR: Workshop and Conference Proceedings*, 2016.
  - H. N. Mhaskar and T. Poggio, Deep vs. shallow networks: An approximation theory perspective, *Analysis and Applications*, 2016.

- **VC Dimension**
  - P. L. Bartlett, N. Harvey, C. Liaw, A. Mehrabian, Nearly-tight VC-dimension and Pseudodimension Bounds for Piecewise Linear Neural Networks, *Journal of Machine Learning Research*, 2019.

- **Optimality results**
  * R. A. DeVore, R. Howard, C. Micchelli, Optimal nonlinear approximation, *manuscripta mathematica*, 1989.
  - D. Yarotsky, Optimal approximation of continuous functions by very deep ReLU networks, *Proceedings of Machine Learning Research*, 2018.
  - D. Yarotsky, A. Zhevnerchuk, The phase diagram of approximation rates for deep neural networks, 2019.
  - J. Lu, Z. Shen, H. Yang, S. Zhang, Deep Network Approximation for Smooth Functions, 2020.

- **Kolmogorov superposition theorem**

* J. Braun, M. Griebel, On a Constructive Proof of Kolmogorov's Superposition Theorem, *Constructive Approximation*, 2009.

– H. Montanelli, H. Yang, Error bounds for deep ReLU networks using the Kolmogorov–Arnold superposition theorem, 2019.

Additional references:

- Allan Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numerica*, 1999.
- M. Anthony, P. L. Bartlett, Neural Network Learning: Theoretical Foundations, 1999.
- N. Cohen, O. Sharir, Y. Levine, R. Tamari, D. Yakira, A. Shashua, Analysis and Design of Convolutional Networks via Hierarchical Tensor Decompositions, 2017.
- N. Cohen, O. Sharir, A. Shashua. On the Expressive Power of Deep Learning: A Tensor Analysis, *JMLR*, 2016.
- M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, J. Sohl-Dickstein, On the Expressive Power of Deep Neural Networks, 2016.
- R. Gribonval, G. Kutyniok, M. Nielsen, F. Voigtlaender, Approximation spaces of deep neural networks., 2019.