Approximation Theory of Neural Networks - List of References

We started our excursion into approximation theory by recalling the universal approximation theorem. We presented a simplified version of Cybenko's original proof. The main references for this result are:

- K. Hornik, M. Stinchcombe, and H. White. *Multilayer feedforward networks are universal approximators.* Neural Netw., 2(5):359–366, 1989. `https://www.sciencedirect.com/science/article/pii/0893608089900208`
- G. Cybenko. *Approximation by superpositions of a sigmoidal function.* Math. Control Signal, 2(4):303–314, 1989. `https://link.springer.com/article/10.1007%2FBF02551274`

The next result we considered was an approximation result that yields arbitrarily good approximations of every continuous function f by neural networks with a finite size. This size is independent of the function f.

- V. Maiorov and A. Pinkus. *Lower bounds for approximation by MLP neural networks.* Neurocomputing, (25):81–91, 1999. `https://www.sciencedirect.com/science/article/pii/S0925231298001118`

The result above uses an overly complicated and completely impractical activation function. For more reasonable activation functions, so-called sigmoidal functions, we recalled an approximation result for Sobolev regular functions. This result was based on a method to emulate B-splines by neural networks.

- H.N Mhaskar and C.A Micchelli. *Approximation by superposition of sigmoidal and radial basis functions.* Adv. in Appl. Math., 13(3):350–373, 1992. `https://www.sciencedirect.com/science/article/pii/019688589290016P`

In applications, mostly the ReLU activation function is used. This function does not fall into the category of sigmoidal functions of the previous results since it is not sufficiently smooth. Approximation results for the ReLU have been analysed in the papers below:

- D. Yarotsky. *Error bounds for approximations with deep ReLU networks.* Neural Netw., 94:103–114, 2017. `https://arxiv.org/pdf/1610.01145.pdf`
- P. Petersen and F. Voigtlaender. *Optimal approximation of piecewise smooth functions using deep ReLU neural networks.*, Neural Netw., in Press. `https://arxiv.org/pdf/1709.05289.pdf`

Finally, we talked about lower bounds on the approximation rates of neural networks. Multiple frameworks can be used to produce such lower bounds. One such framework is that of continuous nonlinear N-width introduced in

- R. DeVore, R. Howard and C. Micchelli, *Optimal non-linear approximation.*, Manuscripta Math., 63(4) 469-478, 1989. `http://www.math.tamu.edu/~rdevore/publications/58.pdf`

Another method is based on the so-called VC-dimension. An excellent introduction to the VC-dimension and estimates of the VC-dimension of sets of neural networks can be found in

If you have any questions, please consider writing an email to `Philipp.Petersen@maths.ox.ac.uk`. If the question has something to do with the material above or approximation theory in general you might even get an informative answer.

- M. Anthony and P.L. Bartlett. *Neural Network Learning: Theoretical Foundations.* Cambridge University Press, 1st edition, 2009.

Results that yield lower bounds from the concepts of VC-dimension and continuous non-linear N-width can, for example, be found in Yarotsky's paper.