

Proposal for SURE 2021

Motivation and Research Question:

Gradient descent is a fundamental algorithm in machine learning, and it finds global minimum in deep neural networks although the loss function may be non-convex. It is believed that over parametrization plays an important role. In previous paper [1], it is proved that gradient descent achieves zero training loss in polynomial time for a deep over-parametrized neural network, but it still remains one important open question:

Can stochastic gradient descent find solutions of zero training loss for deep neural network?

This question is motivating because SGD is heavily used in machine learning algorithms to save computational time. Thus, we hope that SGD also enjoys this nice property.

Bibliography:

[1] S. Du, J. Lee, H. Li, L. Wang, and X. Zhai. Gradient descent finds global minima of deep neural networks. In *International Conference on Machine Learning*, pages 1675–1685. PMLR, 2019.

[2] S. Du, X. Zhai, B. Póczos, A. Singh. Gradient Descent Provably Optimizes Over-parameterized Neural Networks, ICLR, 2019.

[3] Allen-Zhu, Z., Li, Y., and Liang, Y. Learning and generalization in overparameterized neural networks, going beyond two layers. *arXiv preprint arXiv:1811.04918*, 2018a.

[4] Y. Li and Y. Liang. Learning overparameterized neural networks via stochastic gradient descent on structured data. *arXiv preprint arXiv:1808.01204*, 2018.

[5] Allen-Zhu, Z., Li, Y., and Song, Z. A convergence theory for deep learning via overparameterization. *arXiv preprint arXiv:1811.03962*, 2018c.

[6] M. Hardt, B. Recht, and Y. Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *International Conference on Machine Learning*, pages 1225–1234. PMLR, 2016.

Plan:

We plan to start with the references, especially the first one since it is highly relevant. We will study the proof techniques (for GD only) and see how they can be generalized to the cases we are interested in, namely SGD. We will narrow down our direction after reading these references, and start the research by reading more papers, running numerical experiment, and trying to provide theoretical proof or at least intuitions behind the results.