船井情報科学振興財団 留学報告書

川口賢司

The following paper was recently accepted and is scheduled to appear in Neural Computation Vol. 31, No. 7, July 2019.

- Kenji Kawaguchi, Jiaoyang Huang and Leslie Pack Kaelbling. (2019). Effect of Depth and Width on Local Minima in Deep Learning. Neural Computation, Vol. 31 No. 7. To appear.

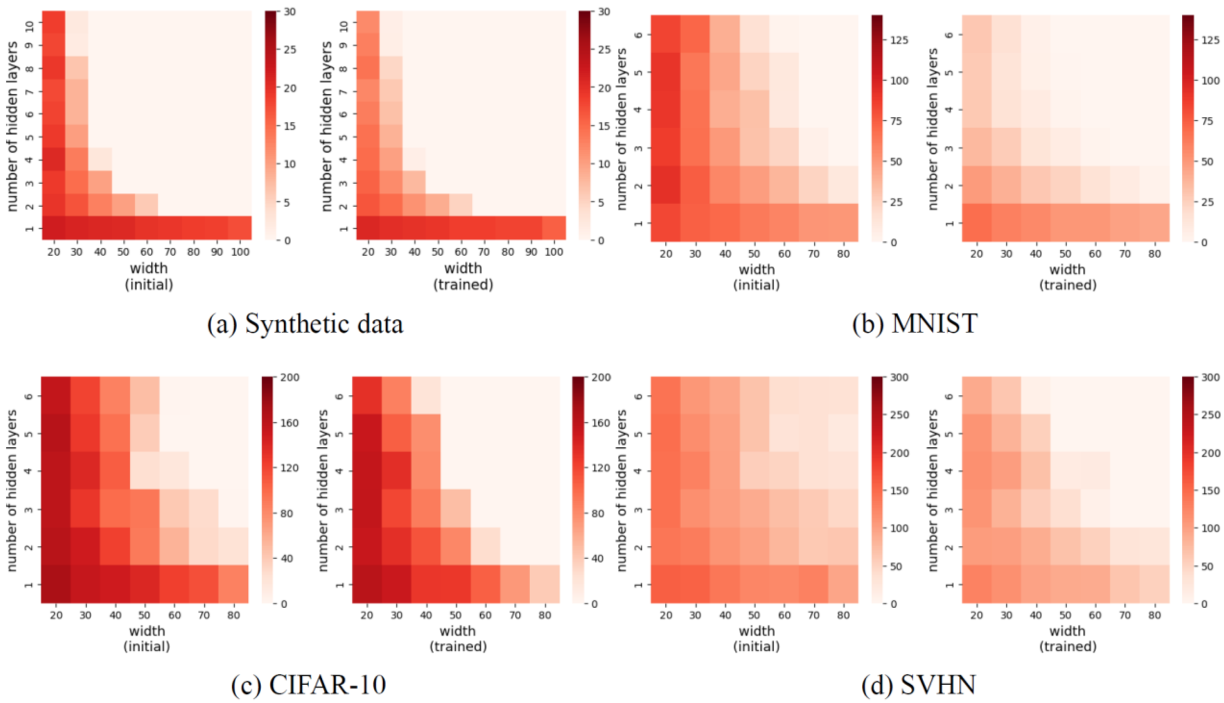

Figure 2 from the paper: it shows that the theoretical guarantee on the value of the loss function at critical points improve as the width and depth increases in various datasets. The x axis is the width, the y axis is the depth, and the heat map represents the value of the loss function.

This paper theoretically analyzes the effects of depth and width on the quality of local minima *without significant over-parameterization and simplification assumptions*. Among other results, this paper theoretically explains the above empirical phenomenon that the quality of critical points and local minima improve towards the global minimum value as depth and width increase, without significant over-parametrization and simplification.

It is important to study this regime without *significant over-parameterization.* A series of recent papers showed that (stochastic) gradient decent can find a global minimum with the assumption of *significant over-parameterization*. However, there is a huge gap between the *significant over-parameterization* assumption in these theoretical papers, and over-parametrization in practice, even though the English word misleadingly suggests otherwise. These theoretical results with significant over-parameterization have a fundamental limitation to explain *representation learning*, and it does not give us a satisfying explanation about practical *deep neural networks* as discussed by the following paper:

- Yehudai, G. and Shamir, O. (2019). On the power and limitations of random features for understanding neural networks. arXiv preprint arXiv:1904.00687.

Roughly speaking, in theoretical studies with significant over-parameterization, there is no effective *representation learning* because a deep neural network simply behaves like a shallow linear model with fixed random features.

Therefore, we want to study the practical regime of deep learning with effective *representation learning* and without the impractical usage of the assumption of significant over-parametrization.

With this in mind, the following paper also analyzes this practical regime without the assumption of significant over-parameterization and model simplification:

- Kenji Kawaguchi and Leslie Pack Kaelbling. (2019). Elimination of all bad local minima in deep learning. arXiv preprint arXiv:1901.00279.

This paper shows that adding a special neuron per output units removes all suboptimal local minima of any deep neural network with an arbitrary loss function, for multi-class classification, binary classification, and regression. Moreover, this paper provides a new proof technique, which gives a new theoretical insight into the elimination of local minima. Furthermore, this paper proves a remaining limitation of eliminating local minima, opening up the need for further research to solve the problem.

For the paper “Elimination of all bad local minima in deep learning”, I was invited to give a talk at a webinar series at PhILMs on 2019/6/3: PhILMs center is a collaboration among Brown University, MIT, Stanford University, University of California Santa Barbara, PNNL, and Sandia National Laboratories.

Beyond deep learning, the following paper provides a novel necessary condition of local minima for general nonconvex machine learning:

- Kenji Kawaguchi and Leslie Pack Kaelbling. (2019). Every Local Minimum is a Global Minimum of an Induced Model. arXiv preprint arXiv:1904.03673.

The proven necessary condition of local minima is also a sufficient condition of global minima for convex machine learning. In particular, for non-convex optimization in machine learning, this paper proves that every local minimum achieves the global optimality of the perturbable gradient basis model at any differentiable point.

I was also invited to give a talk at ICIAM 2019 Minisymposium on Theoretical Foundations of Deep Learning. Other scheduled speakers in the same session are Professor Francis Bach, Professor Rene Vidal, and Dr. Philipp Petersen. I am looking forward to listening talks by all.

船井情報科学振興財団には心より感謝しております。本当にありがとうございま す。