船井情報科学振興財団 留学報告書

川口賢司

Since the last report, several new papers have been published, including:

[1] Kenji Kawaguchi and Jiaoyang Huang. Gradient Descent Finds Global Minima for Generalizable Deep Neural Networks of Practical Sizes. In Proceedings of the 57th Allerton Conference on Communication, Control, and Computing (Allerton), 2019.

[2] Kenji Kawaguchi and Yoshua Bengio. Depth with Nonlinearity Creates No Bad Local Minima in ResNets. Neural Networks, 118, 167-174, 2019.

[3] Kenji Kawaguchi, Jiaoyang Huang and Leslie Pack Kaelbling. Every Local Minimum Value is the Global Minimum Value of Induced Model in Non-convex Machine Learning. Neural Computation, 31(12), 2293-2323, MIT press, 2019.

[4] Ameya D. Jagtap, Kenji Kawaguchi, George E. Karniadakis. Adaptive Activation Functions Accelerate Convergence in Deep and Physics-informed Neural Networks. Journal of Computational Physics, accepted, 2019.

The first paper [1] first provides experimental observations of the gap between the current theory and practice, and then proves novel theorems in order to close the gap. Namely, the paper first provides the observation in Figure 1, and then theoretically proves that for deep neural networks, gradient decent can find a global minimum with the practical degree of over-parameterization that matches with the degree of over-parameterization used in practice. It also proves both the upper and lower bound on the degree of over-parametrization required to guarantee a global minimum for deep neural networks.

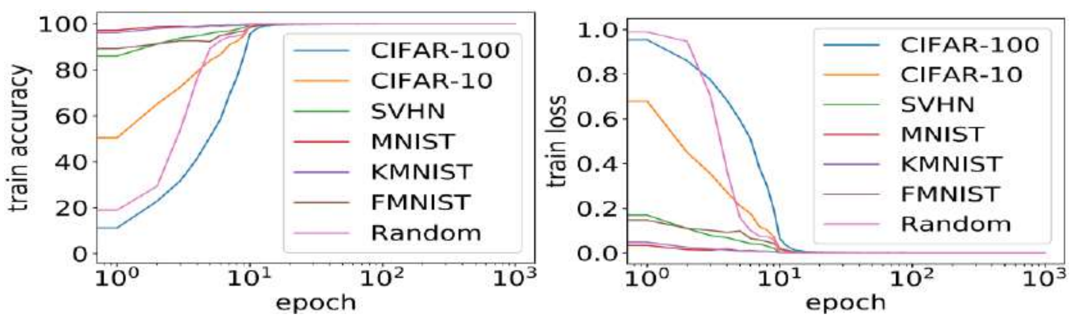

Figure 1 from paper [1]: it shows that deep neural networks in practice can achieve 100% test accuracy (and approximately zero training loss) for all datasets without the extreme degree of over-parameterization that is required by the previous theory, but not required by the proposed theory.

Whereas the first paper [1] requires over-parameterization, the second paper [2] does not require over-parametrization and still theoretically guarantees the quality of local minima for nonconvex optimization of deep ResNets. In optimization, without convexity (and invexity), we have the concern of optimization methods getting struck around a region of a local minimum (Figure 1). In machine learning, this concern occurs when training deep neural networks via nonconvex (and non-invex) optimization. The second paper [2] shows that we can mitigate this concern by using deep ResNets, instead of fully-connected deep neural networks.

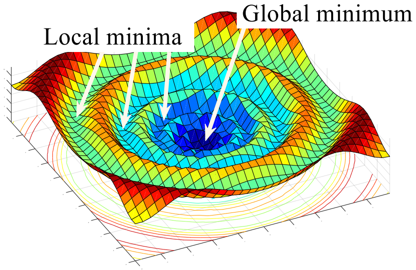

Figure 1. Illustration of local minima in non-convex optimization landscapes

Whereas the first and second papers [1]-[2] focus on the specific structures of neural networks (over-parameterization or ResNets), the third paper [3] provides a novel theory and geometric viewpoints for general nonconvex machine learning models, including all typical deep neural networks. For stationary points and local minima, the theory in the third paper [3] concludes global optimality in the data-dependent and architecture-dependent hypothesis space, which also entails the previous results for particular architectures such as over-parametrized networks and ResNets.

先月 Toyota Technological Institute at Chicago に訪問してきました。研究面では、新しい研究課題を教えていただく機会や他分野に触れる機会があり大変勉強になりました。自分の得意とする分野が定まってきているためか、最近他大学への訪問中や MIT 内で生産的な議論をする機会が増えている気がします。現在これらの議論が共同研究につながり結果が少しずつ出始めています。来年中に共同研究の結果もご報告できるよう頑張ります。

ところで、今までランニングを中心に運動していたのですが、ボストンの寒さに負けて運動しない日々が続くことが多々ありました。運動不足を改善すべく室内で運動できるスカッシュを 4 か月前くらいから始めました。スカッシュのおかげで運動不足が改善され、日々充実している気がします。寒い地域に長期留学し運動不足を感じたときは、スカッシュなどの室内競技を試しみることをお勧めします。

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。