Journal of Machine Learning

Embedding Principle: a hierarchical structure of loss landscape of deep neural networks

Yaoyu Zhang, Yuqing Li, Zhongwang Zhang, Tao Luo, Zhi-Qin John Xu

Summary for general readers:

Understanding the loss landscape of a deep neural network is obviously important in analyzing the training trajectory and the generalization performance. This work proposes a new approaching for studying this problem by examining the (embedding) relation between the loss landscapes of neural networks of different widths. The paper proves a general Embedding Principle, namely the loss landscape of a neural network "contains" all critical points of the landscape for narrower neural networks. The paper also demonstrates that (i) critical points embedded from narrower neural networks form submanifolds; (ii) a local minimum is more likely to become a strict saddle point in the landscape of wider neural networks but not vice versa.