World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:14, No:06, 2020

Convergence Analysis of Training Two-Hidden-Layer Partially Over-Parameterized ReLU Networks via Gradient Descent

Authors : Zhifeng Kong

Abstract : Over-parameterized neural networks have attracted a great deal of attention in recent deep learning theory research, as they challenge the classic perspective of over-fitting when the model has excessive parameters and have gained empirical success in various settings. While a number of theoretical works have been presented to demystify properties of such models, the convergence properties of such models are still far from being thoroughly understood. In this work, we study the convergence properties of training two-hidden-layer partially over-parameterized fully connected networks with the Rectified Linear Unit activation via gradient descent. To our knowledge, this is the first theoretical work to understand convergence properties of deep over-parameterized networks without the equally-wide-hidden-layer assumption and other unrealistic assumptions. We provide a probabilistic lower bound of the widths of hidden layers and proved linear convergence rate of gradient descent. We also conducted experiments on synthetic and real-world datasets to validate our theory.

Keywords : over-parameterization, rectified linear units ReLU, convergence, gradient descent, neural networks
Conference Title : ICNIP 2020 : International Conference on Neural Information Processing
Conference Location : San Francisco, United States
Conference Dates : June 05-06, 2020