Constructive Universal Approximation and Sure Convergence for Multi-Layer Neural Networks

Chien-Ming Chi

Institute of Statistical Science, Academia Sinica

Abstract

We propose o1Neuro, a new neural network model built on sparse indicator activation neurons, with two key statistical properties. (1) Constructive universal approximation: At the population level, a deep o1Neuro can approximate any measurable function of $\boldsymbol{X}$, while a shallow o1Neuro suffices for additive models with two-way interaction components, including XOR and univariate terms, assuming $\boldsymbol{X} \in [0,1]^p$ has bounded density. Combined with prior work showing that a single-hidden-layer non-sparse network is a universal approximator, this highlights a trade-off between activation sparsity and network depth in approximation capability. (2) Sure convergence: At the sample level, o1Neuro's optimization reaches an optimal model with probability approaching one after sufficiently many update rounds, and we provide an example showing that the required number of updates is well bounded under linear data-generating models. Empirically, o1Neuro is compared with XGBoost, Random Forests, and TabNet for learning complex regression functions with interactions, demonstrating superior predictive performance on several benchmark datasets from OpenML and the UCI Machine Learning Repository with $n = 10{,}000$, as well as on synthetic datasets with $100 \le n \le 20{,}000$.

Keywords: greedy approximation, boosting machine, activation sparsity, idle neurons, nonconvex optimization problems