Analyzing Tree Architectures in Ensembles via Neural Tangent Kernel

Ryuichi Kanoh¹,², Mahito Sugiyama¹,²
¹National Institute of Informatics, ²The Graduate University for Advanced Studies, SOKENDAI

The 11th International Conference on Learning Representations (ICLR 2023), May 1-5, 2023

Contribution

- We formulate and analyze the Neural Tangent Kernel (NTK) induced by soft tree ensembles for arbitrary tree architectures

Soft Tree Ensemble

- A variant of trees that inherits characteristics of neural networks
  - Splitting rules and leaf values are updated with gradient descent
  - Unlike typical decision trees, feature engineering is included in training

- The NTK for ensembles of perfect binary trees is known
  - It converges to a closed kernel when we consider infinite trees ($M \to \infty$)

NTK for Arbitrary Tree Architectures

- The NTK induced by infinite ensembles of arbitrary trees can be decomposed by the NTKs induced by rule ensembles
  - Characterized only by the number of tree leaves per depth

NTK for infinite ensembles of rule sets

$\Theta^{(D, \text{Rule})}(x_i, x_j) := \lim_{M \to \infty} \tilde{\Theta}^{(D, \text{Rule})}(x_i, x_j)$

$= D \sum_{d=1}^{D} \tilde{\Theta}^{(d, \text{Rule})(\text{internal nodes})}(x_i, x_j) + \tilde{\Theta}^{(d, \text{Rule})(\text{leaves})}(x_i, x_j)$

NTK for infinite ensembles of arbitrary trees

$\Theta^{(A, \text{Tree})}(x_i, x_j) := \lim_{M \to \infty} \tilde{\Theta}^{(A, \text{Tree})}(x_i, x_j)$

$= \sum_{d=1}^{Q(d)} \Theta^{(A, \text{Tree})}(x_i, x_j)$

Q(d): The number of leaves per depth at d

Tree Equivalence

- Training behavior can be equivalent even for non-isomorphic trees
  - The NTK depends only on the number of leaves per depth

Output dynamics for test data points. Each line color corresponds to each data point

Case Study: Decision List

- The NTK for infinitely deep decision lists does not degenerate
  - This is in contrast to the perfect binary trees

Decision list: a binary tree that grows in only one direction

- The degeneracy leads to worse generalization performance
  - Unable to distinguish between a 90-degree and a 180-degree difference

Reference

[Neural Tangent Kernel]
- Jacot et al. (NeurIPS 2018), Neural Tangent Kernel: Convergence and Generalization in Neural Networks
- Kanoh & Sugiyama (ICLR 2022), A Neural Tangent Kernel Perspective of Infinite Tree Ensembles

[Soft Tree]
- Kontschieder et al. (ICCV 2015), Deep Neural Decision Forests
- Popov et al. (ICLR 2020), Neural Oblivious Decision Ensembles for Deep Learning on Tabular Data
- Hazimeh et al. (ICML 2020), The Tree Ensemble Layer: Differentiability meets Conditional Computation