A Neural Tangent Kernel Perspective of Infinite Tree Ensembles

Ryuichi Kanoh[1,2], **Mahito Sugiyama**[1,2]
[1]National Institute of Informatics, [2]The Graduate University for Advanced Studies, SOKENDAI

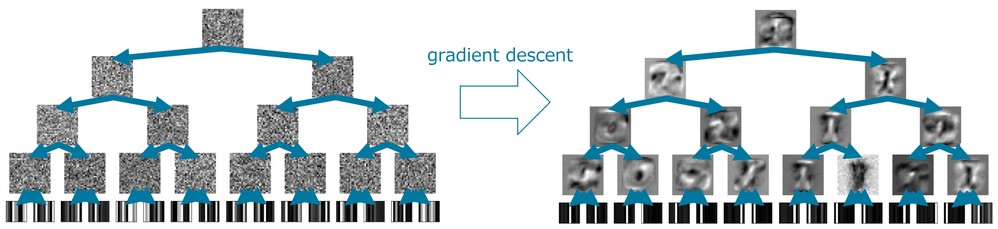
The 10th International Conference on Learning Representations (ICLR 2022), April 25-29, 2022

Our Contribution

- **First extension of the NTK concept to the tree ensemble models**
  - Tree Neural Tangent Kernel: TNTK
  - Derive the analytical form for the TNTK induced by infinitely many trees
- **Theoretical support for empirical techniques**
  - Parameter sharing, depth tuning, splitting operation modification, ...

Soft Tree Ensemble

- **A variant of tree ensembles 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 the training



Neural Tangent Kernel

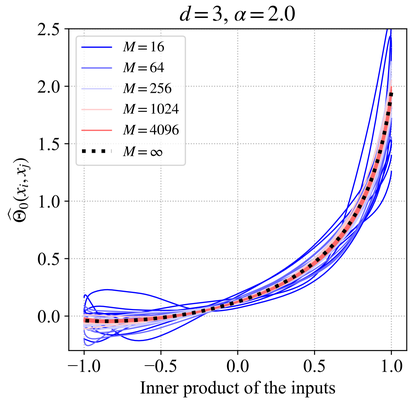
- **A Kernel that describes the evolution of the models during their training**
  - Training behavior can be analyzed using kernel methods
- **Many successes for infinitely wide neural networks**
  - CNN, RNN, GNN, ...
  - We extend the NTK concept to infinitely many soft trees!!

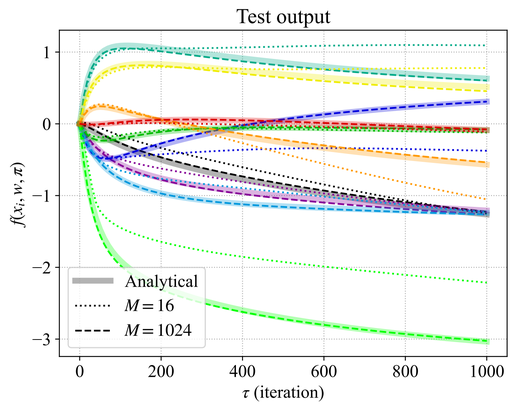
Properties of the TNTK

- **The TNTK induced by infinite trees converges to a deterministic kernel**

$$\lim_{M\to\infty} \widehat{\Theta}_0^{(d)}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \underbrace{2^d d\, \Sigma(\boldsymbol{x}_i, \boldsymbol{x}_j)(\mathcal{T}(\boldsymbol{x}_i, \boldsymbol{x}_j))^{d-1}\dot{\mathcal{T}}(\boldsymbol{x}_i, \boldsymbol{x}_j)}_{\text{contribution from inner nodes}} + \underbrace{(2\mathcal{T}(\boldsymbol{x}_i, \boldsymbol{x}_j))^d}_{\text{contribution from leaves}}$$

- **As the #trees increases, the change from the initial value becomes smaller**
  - Training behavior of infinite trees is analytically tractable





Implications for Practical Techniques

- **Infinite oblivious trees induce the same TNTK compared with normal trees**
  - Parameter sharing does not degrade performance
- **Overly deep trees induce degenerated kernel**
  - Support for poor generalization performance of too deep trees