For favour of Posting

THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

PHD STUDENT SEMINAR

Approximation Bounds for Transformer Networks with Application to Regression

By

Mr. Bokai YAN

Abstract

We explore the approximation capabilities of Transformer networks for H\"older and Sobolev functions, and apply these results to address nonparametric regression estimation with dependent observations. First, we establish novel upper bounds for standard Transformer networks approximating sequence-to-sequence mappings whose component functions are H\"older continuous with smoothness index $\gamma \in (0,1]$. To achieve an approximation error ε under the L^p-norm for $p \in [1, \infty]$, it suffices to use a fixed-depth Transformer network whose total number of parameters scales as $\varepsilon^{-d_x n / \gamma}$. This result not only extends existing findings to include the case $p = \infty$, but also matches the best known upper bounds on number of parameters previously obtained for fixed-depth FNNs and RNNs. Similar bounds are also derived for Sobolev functions. Second, we derive explicit convergence rates for the nonparametric regression problem under various β-mixing data assumptions, which allow the dependence between observations to weaken over time. Our bounds on the sample complexity impose no constraints on weight magnitudes. Lastly, we propose a novel proof strategy to establish approximation bounds, inspired by the Kolmogorov-Arnold representation theorem. We show that if the self-attention layer in a Transformer can perform column averaging, the network can approximate sequence-to-sequence H\"older functions, offering new insights into the interpretability of self-attention mechanisms.

Date : 12 May 2025, Monday

Time : 12:00noon

Venue : Room 2612B (Lifts 31-32)

All are Welcome!