The Hong Kong University of Science and Technology

Department of Mathematics

PhD THESIS EXAMINATION

Theoretical Study of Neural Network Models: Error Bounds on Approximation, Generalization, and Optimization with Applications to Regression and Partial Differential Equations

By

Mr. Yanming LAI

ABSTRACT

The remarkable success of machine learning in computer vision, natural language processing, and related domains has spurred vigorous development of its theoretical foundations. Neural networks are a core model of machine learning, achieving complex function approximation and automatic feature learning by simulating the connections between neurons in the human brain. As a nonparametric estimation method, the error analysis of machine learning typically comprises three fundamental components: approximation error, generalization error, and optimization error. This thesis presents a comprehensive investigation of these three error types for neural networks, along with convergence rate analysis when applied to solving both regression problems and partial differential equations (PDEs). In the first part of this thesis, we derive the convergence rate for solving regression problems using three-layer logistic overparameterized feedforward neural networks (FNNs) trained with gradient descent (GD). In the second part, within the framework of the Deep Ritz Method (DRM), we derive the convergence rate for solving second-order elliptic equations with three different types of boundary conditions using three-layer overparameterized tanh FNNs trained with projected gradient descent (PGD). In both parts, following the tradition of nonparametric estimation, our error bounds are expressed in terms of the sample size n. Our results also provide a quantitative description of various parameters, including the depth and width of the neural network, the training step size and number of iterations of the optimization algorithm. In the third part of this thesis, we focus on the expressive power (approximation capability) of the Transformer model, which has recently demonstrated strong performance in large language models but remain theoretically underexplored compared to classical FNNs with extensive literature. Specifically, we investigate the approximation of the Hölder continuous function class by Transformers and construct several Transformers that can overcome the curse of dimensionality. These results demonstrate that Transformers possess super expressive power.

Date : 16 July 2025, Wednesday

Time : 10:00 am

Venue : Room 5506 (Lifts 25/26)

***Thesis Examination Committee*:**

Chairman	**:**	**Prof. Xiangru ZHANG, CIVL/HKUST**
Thesis Supervisor	**:**	**Prof. Yang WANG, MATH/HKUST**
Member	**:**	**Prof. Yang XIANG, MATH/HKUST**
Member	**:**	**Prof. Xinzhou GUO, MATH/HKUST**
Member	**:**	**Prof. Ding PAN, PHYS/HKUST**
External Examiner	**:**	**Prof. Bangti JIN, Department of Mathematics/ The Chinese University of Hong Kong** *(via online mode)*

(Open to all faculty and students)

The student's thesis is now being displayed on the reception counter in the General Administration Office (Room 3461).