The Hong Kong University of Science and Technology

Department of Mathematics

MPhil THESIS EXAMINATION

The Impacts of Over-parameterization under Distribution Shift

By

Miss Yifan HAO

ABSTRACT

In recent years, machine learning models have achieved success based on the independently and identically distributed (IID) assumption. However, this assumption can be easily violated in real-world applications, leading to both adversarial attack and the Out-of-Distribution (OOD) problem. Understanding how modern over-parameterized models behave under non-trivial distributional shifts is essential, as current theoretical understanding is insufficient. Here we investigate the performances of overparameterized model in terms of both adversarial robustness and OOD generalization under the commonly-adopted "benign overfitting" conditions. Our results contains two parts: (i) for adversarial robustness, we prove that even if the ground truth itself is robust to adversarial examples, and the benignly overfitted model is benign in terms of the "standard" out-of-sample risk objective, this benign overfitting process can be harmful when out-of-sample data are subject to adversarial manipulation; (ii) for OOD generalization, we focus on a scenario where the benign overfitting estimators demonstrate a constant excess OOD loss, despite achieving zero excess in-distribution (ID) loss, and prove that further increasing the model's parameterization can significantly reduce the OOD loss; (iii) we further show that model ensembles can also enhance the OOD testing loss, achieving a similar effect to increasing the model capacity. These results offer theoretical insights into the intriguing empirical phenomenon in "benign overfitting" regime.

Date : 3 June 2024, Monday
Time : 4:30 p.m.
Venue : Room 2463 (Lifts 25-26)

Thesis Examination Committee

Chairman	**:**	**Prof. Xinzhou GUO, MATH /HKUST**
Thesis Supervisor	**:**	**Prof. Kani CHEN, MATH/HKUST**
Member	**:**	**Prof. Yuan YAO, MATH/HKUST**

(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).