*For favour of Posting*

**THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY**

**Department of Mathematics**

# SEMINAR ON APPLIED MATHEMATICS

## Towards unlocking the mystery of adversarial fragility of neural networks

**By**


**Prof. Weiyu Xu**

University of Iowa



**Abstract**

In this talk, we study the adversarial robustness of deep neural networks for classification tasks. We look at the smallest magnitude of possible additive perturbations that can change the output of a classification algorithm. We provide a matrix-theoretic explanation of the adversarial fragility of deep neural networks for classification. In particular, our theoretical results show that a neural network's adversarial robustness can degrade as the input dimension d increases. Analytically we show that neural networks' adversarial robustness can be only 1 over the square root of d fraction of the best possible adversarial robustness. Our matrix-theoretic explanation is consistent with an earlier information-theoretic feature-compression-based explanation for the adversarial fragility of neural networks.


**Date : 26 July 2024 (Friday)**

**Time : 10:00a.m.-11:00a.m.**

**Venue : Room 1409 (near Lift 25/26)**

*All are Welcome!*