数学与系统科学研究院

计算数学所学术报告

报告人：　　朱占星 博士

（ 北京大学数学学院 ）

报告题目：

Adversarial Training for Deep Learning: A General Framework for Improving Robustness and Generalization

邀请人：　明平兵 研究员

报告时间：**2019** 年 **9** 月 **19** 日（周四）

上午 **10:00-11:00**

报告地点：　数学院南楼二层

202 教室

Abstract:

Deep learning has achieved tremendous success in various application areas, such as computer vision, natural language processing, game playing (AlphaGo), etc. Unfortunately, recent works show that an adversary is able to fool the deep learning models into producing incorrect predictions by manipulating the inputs maliciously. The corresponding manipulated samples are called adversarial examples. This vulnerability issue dramatically hinders the deployment of deep learning, particularly in safety-critical applications.

In this talk, I will introduce various approaches for how to construct adversarial examples. Then I will present a framework, named as adversarial training, for improving robustness of deep networks to defense the adversarial examples, and how to accelerate the training. Moreover,I will show that the introduced adversarial learning framework can be extended as an effective regularization strategy to improve the generalization in semi-supervised learning.

欢迎大家参加！