中央研究院統計科學研究所

學 術 演 講

講 題：Adversarial Classification

演 講 人：Prof. Fabrizio Ruggeri
Institute for Applied Mathematics and Information Technologies, National Research Council of Italy, Italy

時 間：2025-04-21 (Mon.) 10:00-12:00

地 點：Auditorium, B1F, Institute of Statistical Science；The tea reception will be held at 10:10.

備 註：Online live streaming through Cisco Webex will be available.

Abstract

In multiple domains such as malware detection, automated driving systems, or fraud detection, classification algorithms are susceptible to being attacked by malicious agents willing to perturb the value of instance covariates in search of certain goals. Such problems pertain to the field of adversarial machine learning and have been mainly dealt with, perhaps implicitly, through game-theoretic ideas with strong underlying common knowledge assumptions. These are not realistic in numerous application domains in relation to security. We present an alternative statistical framework that accounts for the lack of knowledge about the attacker's behavior using adversarial risk analysis concepts.

中 央 研 究 院

統計科學研究所