統計科學研究所
INSTITUTE OF STATISTICAL SCIENCE

統計所學術演講

中研院統計所

學　術　演　講

講　題：Data Perturbation

講　者：Prof. Xiaotong Shen (沈曉彤 教授)

(School of Statistics, University of Minnesota, USA)

時　間：2022年11月14日(星期一)，11:00-12:00

地　點：統計所B1演講廳

Abstract

Data perturbation is a technique for generating synthetic data by adding "noise" to original data, which has a wide range of applications, primarily in data security. Yet, it has not received much attention within data science. In this presentation, I will describe a fundamental principle of data perturbation that preserves the distributional information, thus ascertaining the validity of the downstream analysis and a machine learning task while protecting data privacy. Applying this principle, we derive a scheme to allow a user to perturb data nonlinearly while meeting the requirements of differential privacy and statistical analysis. It yields credible statistical analysis and high predictive accuracy of a machine learning task. Finally, I will highlight multiple facets of data perturbation through examples.

※ 英文演講，實體與線上視訊同步進行。