



統計科學研究所

INSTITUTE OF
STATISTICAL SCIENCE



S E M I N A R



S T A T A S

Seminar

Title : Data Perturbation

Speaker : Prof. Xiaotong Shen (沈曉彤 教授)
(School of Statistics, University of Minnesota, USA)

Time : 11:00 AM~12:00 PM, Monday, Nov 14, 2022

Place : Auditorium, B1F, Institute of Statistical Science

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.

※ Lecture in English. Online live streaming through Cisco Webex will be available.