

淡江大學數學系演講公告

主講人：黃名鉞 副研究員(中研院統計科學研究所)

題目：Efficient Data Integration Under Prior Probability Shift

日期：112年10月17日(星期二)

時間：下午13:10開始

地點：科學館 S433

摘要 Abstract:

Conventional supervised learning usually operates under the premise that data are collected from a homogeneous underlying population. However, challenges may arise when integrating new data from different populations, resulting in a phenomenon known as dataset shift. This paper focuses on prior probability shift, a specific form of dataset shift, where the distribution of the outcome varies across different datasets but the conditional distribution of features given the outcome remains the same. To tackle the challenges posed by this shift, we propose a maximum likelihood estimation method that efficiently amalgamates information from multiple sources under prior probability shift. Unlike existing methods that are restricted to discrete outcomes, the proposed approach accommodates both discrete and continuous outcomes. It also handles high-dimensional covariate vectors through variable selection using an adaptive LASSO penalty, producing efficient estimates that possess the oracle property. Moreover, a novel semiparametric likelihood ratio test is proposed to check the validity of prior probability shift assumptions by embedding the null conditional density function into Neyman's smooth alternatives and testing study-specific parameters. We demonstrate the effectiveness of our proposed method through extensive simulations and two real data examples. The proposed methods serve as a useful addition to the repertoire of tools for addressing challenges that arise from dataset shifts in machine learning. This is a joint work with Dr. Jing Qin in National Institute of Allergy and Infectious Diseases, National Institutes of Health and Prof. Chiung-Yu Huang in University of California, San Francisco.

備註：本業務與聯合國永續發展目標 SDG4 優質教育連結

AJ+SDGs = ∞

歡迎參加

淡江大學數學系 敬邀