School of Data Science
香港城市大學
City University of Hong Kong

Hong Kong Institute for Data Science
香港城市大學
City University of Hong Kong

Classification with Imperfect Training Labels

Date: 18 November 2020 (Wednesday)

Time: 11:30am – 12:30pm

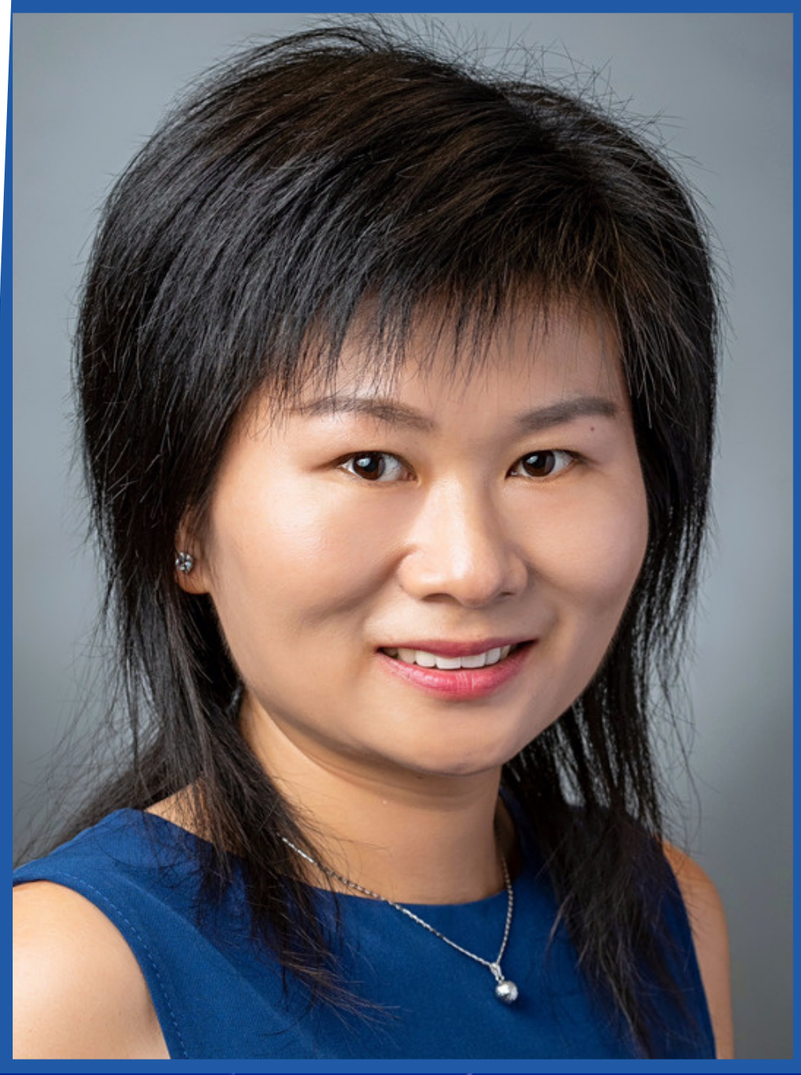
Seminar link: https://cityu.zoom.us/j/99661456506

ABSTRACT

We study the effect of imperfect training data labels on the performance of classification methods. In a general setting, where the probability that an observation in the training dataset is mislabelled may depend on both the feature vector and the true label, we bound the excess risk of an arbitrary classifier trained with imperfect labels in terms of its excess risk for predicting a noisy label. This reveals conditions under which a classifier trained with imperfect labels remains consistent for classifying uncorrupted test data points. Furthermore, under stronger conditions, we derive detailed asymptotic properties for the popular k-nearest neighbour, support vector machine and linear discriminant analysis classifiers. One consequence of these results is that the k-nearest neighbour and support vector machine classifiers are robust to imperfect training labels, in the sense that the rate of convergence of the excess risk of these classifiers remains unchanged; in fact, our theoretical and empirical results even show that in some cases, imperfect labels may improve the performance of these methods. The linear discriminant analysis classifier is shown to be typically inconsistent in the presence of label noise unless the prior probabilities of the classes are equal. Our theoretical results are supported by a simulation study. This is a joint work with Tim Cannings and Richard Samworth.

Enquiries: hkids@cityu.edu.hk

ONLINE SEMINAR



Prof Yingying FAN

GUEST SPEAKER'S PROFILE

Prof Yingying FAN is Professor and Dean's Associate Professor in Business Administration in Data Sciences and Operations Department of the Marshall School of Business at the University of Southern California, Professor in Departments of Economics and Computer Science at USC, and an Associate Fellow of USC Dornsife Institute for New Economic Thinking (INET). She received her Ph.D. in Operations Research and Financial Engineering from Princeton University in 2007. She was Lecturer in the Department of Statistics at Harvard University from 2007-2008. Her research interests include statistics, data science, machine learning, economics, big data and business applications, and artificial intelligence and blockchain. Her latest works have focused on statistical inference for networks, texts, and AI models empowered by some most recent developments in random matrix theory and statistical learning theory.

Her papers have been published in journals in statistics, economics, computer science, information theory, and biology. She is the recipient of Fellow of Institute of Mathematical Statistics (2020), Fellow of American Statistical Association (2019), NIH R01 Grant (2018), the Royal Statistical Society Guy Medal in Bronze (2017), USC Marshall Dean's Award for Research Excellence (2017), the USC Marshall Inaugural Dr. Douglas Basil Award for Junior Business Faculty (2014), the American Statistical Association Noether Young Scholar Award (2013), NSF Faculty Early Career Development (CAREER) Award (2012), Zumberge Individual Award from USC's James H. Zumberge Faculty Research and Innovation Fund (2010), USC Marshall Dean's Award for Research Excellence (2010), and NSF Grant (2009), as well as a Plenary Speaker at the 2011 Institute of Mathematical Statistics Workshop on Finance, Probability, and Statistics held at Columbia University. She has served as an associate editor of Journal of the American Statistical Association (2014-present), Journal of Econometrics (2015-2018), Journal of Business & Economic Statistics (2018-present), The Econometrics Journal (2012-present), and Journal of Multivariate Analysis (2013-2016), as well as on the Institute of Mathematical Statistics Committee for the Peter Hall Early Career Prize (2020-2023).

All are welcome