Statistics Seminar

Department of Mathematical Sciences

DATE:	Thursday, May 5, 2022
TIME:	1:15pm – 2:15pm
LOCATION:	Zoom meeting
SPEAKER:	Zhou Wang, Binghamton University
TITLE:	Learning Acceptance Regions for Many Classes with Anomaly Detection

Abstract

Consider a hypothesis test that determines if an observation belongs to a particular class in a multicategory classification setting. The set of all observations for which the null hypothesis is accepted is called the acceptance region. Set-valued classification, a new classification paradigm that aims to identify all the plausible classes that an observation may belong to, can be obtained by constructing acceptance regions for all classes. There is a growing literature on set-valued classification; however, many existing methods do not take into account the possibility that a new class that never appeared in the training data suddenly appears in the test data. Moreover, many existing methods are based on complex numerical optimization that aims to train a set of decision functions for all the classes simultaneously, which can add a significant computational burden especially when the number of classes is large. In this article, we propose a Generalized Prediction Set (GPS) approach to estimate the acceptance region for each class, while taking into account the possibility of a new class in the test data. The proposed classifier minimizes the expected size of the prediction set for each observation while guaranteeing that the class-specific accuracy is at least a pre-specified value. Unlike previous methods, the proposed method achieves a good balance between accuracy, efficiency, and anomaly detection rate. Moreover, our method can be applied in parallel to all the classes so as to alleviate the computational burden. Both theoretical analysis and numerical experiments are conducted to illustrate the effectiveness of the proposed method.