Statistics Seminar

Department of Mathematical Sciences

DATE:	Thursday, September 26, 2019
TIME:	1:15pm – 2:15pm
LOCATION:	WH 100E
SPEAKER:	Zhou Wang, Binghamton University
TITLE:	Cautious Deep Learning with Conformal Prediction

Abstract

Most classifiers operate by selecting the maximum of an estimate of the posterior probability p(y|x) where x stands for the features of the instance to be classified and y denotes its label. This often results in a hubristic bias: overconfidence in the assignment of a definite label. Usually, observations are concentrated on a small volume but the classifier provides definite predictions for the entire space. We propose constructing conformal prediction sets which contain a set of labels rather than a single label with probability $1 - \alpha$ containing the true label. The construction based on p(x|y) rather than p(y|x) results in a cautious classifier: it outputs the null set – meaning "I don't know" – when the object does not resemble the training examples. An important property of our approach is that adversarial attacks are likely to be predicted as the null set or would also include the true label. We demonstrate the performance on the ImageNet dataset and the CelebA and IMDB-Wiki facial datasets using high dimensional features obtained from state of the art convolutional neural networks.