מכון ויצמן למדע

THE WEIZMANN INSTITUTE OF SCIENCE

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Machine Learning and Statistics Seminar

Room 1 ,Ziskind Building
on Wednesday, Nov 28, 2018
at 11:15

Michael Kim
Stanford

Fairness through Computationally-Bounded Awareness

Abstract:

As algorithmic prediction systems have become more widespread, so too have concerns that these systems may be discriminatory against groups of people protected by laws and ethics. We present a recent line of work that takes a complexity theoretic perspective towards combating discrimination in prediction systems. We'll focus on fair classification within the versatile framework of Dwork et al. [ITCS'12], which assumes the existence of a metric that measures similarity between pairs of individuals. Unlike earlier work, we do not assume that the entire metric is known to the learning algorithm; instead, the learner can query this metric a bounded number of times. We propose a new notion of fairness called *metric multifairness* and show how to achieve this notion in our setting. Metric multifairness is parameterized by a similarity metric d on pairs of individuals to classify and a rich collection C of (possibly overlapping) "comparison sets" over pairs of individuals. At a high level, metric multifairness guarantees that *similar subpopulations are treated similarly*, as long as these subpopulations are identified within the class C.