

Short course information:

Fairness in Data Science: Criteria, Algorithms, and Open Problems

Systematic biases present in our society influence the way data is collected and stored, the way variables are defined, and the way scientific findings are put into practice as policy. Automated decision procedures and learning algorithms applied to such data may serve to perpetuate existing injustice or unfairness in our society. Increasing commoditization of statistical and machine learning methods led to a series highly publicized instances of learning algorithms producing inappropriate, discriminatory, or otherwise harmful outputs. As a response, a flurry of research activity aimed to quantitatively describe various aspects of fairness and bias in data science, as well as develop new approaches to learning and estimation from data that takes fairness criteria into account. In this short course, we will review a variety of fairness criteria that have been developed, along with algorithms that aim to be ‘fairness-aware’ in various ways, with a particular emphasis on methods rooted in causal inference. We will conclude by describing a variety of methodological and translational problems that remain in this rapidly growing subfield of data science.