Title: Algorithmic fairness

Organizers: Sam Corbett-Davies, Rachel Cummings, Sharad Goel, Jaime Morgenstern

Description: There is increasing concern that decision making tools informed by supervised learning might exacerbate societal biases. This has produced an explosion of competing definitions of what it means for an algorithm to be fair. The most popular definitions consider error rate parity (e.g., that false positive or false negative rates should be equal across groups), calibration (that outcomes should be independent of group membership conditional on predictions), and anti-classification (that protected characteristics should not be used at decision time).

This tutorial will begin by describing these definitions, before demonstrating the subtle problems with each concept as a fairness objective. In doing so, we draw on classical and modern ideas from statistics, economics, and legal theory. By articulating problems in the foundation of this nascent subfield of algorithmic fairness, we hope to offer researchers and practitioners a way to advance the area. The first part of the tutorial will conclude with a discussion of other common fairness concerns -- including the potential for bias in the data -- and some outstanding challenges to implementing fair algorithms in practice.

In the second part of the tutorial, we give a brief overview of the literature on economic models of discriminatory behavior. Some of these issues stem from the ability to price-match and others from the observability of the protected attribute when combined with imperfect signalling. We add to this list the possibility that agents might not be risk-neutral, causing them to favor lower-variance models over models which might have higher accuracy but less predictable behavior. For example, training one classifier on men and another on women may improve accuracy but will also reduce the effective training set size for each model, therefore increasing the variance of the models' performance out of sample.

Format: 3 hours total presentation time, broken into two 1.5 hour blocks with time for questions and extended discussion.