Algorithmic Design: Fairness Versus Accuracy

Annie LIANG, Jay LU, and Xiaosheng MU

Abstract :
Algorithms are increasingly used to guide consequential decisions, such as who should be granted bail or be approved for a loan. Motivated by growing empirical evidence, regulators are concerned about the possibility that the errors of these algorithms differ sharply across subgroups of the population. What are the tradeoffs between accuracy and fairness, and how do these tradeoffs depend on the inputs to the algorithm? We propose a model in which a designer chooses an algorithm that maps observed inputs into decisions, and introduce a fairness-accuracy Pareto frontier. We identify how the algorithm's inputs govern the shape of this frontier, showing (for example) that access to group identity reduces the error for the worse-off group everywhere along the frontier. We then apply these results to study an ``input-design problem where the designer controls the algorithm's inputs (for example, by legally banning an input), but the algorithm itself is chosen by another agent. We show that: (1) all designers strictly prefer to allow group identity if and only if the algorithm's other inputs satisfy a condition we call group-balance; (2) all designers strictly prefer to allow any input (including potentially biased inputs such as test scores) so long as group identity is permitted as an input, but may prefer to ban it when group identity is not.