Title: Towards Actionable Explanations from Black-Box Decision-Making Systems.

Abstract:
Automated systems are increasingly applied to individualized decision making. Thus, an individual can suffer an undesirable outcome (e.g. denied credit) irrespective of whether the decision is fair or accurate. We consider the task of providing an actionable set of changes which an individual can undertake in order to improve the automated decision. Our approach models the underlying data distribution or manifold. We then provide a mechanism to generate the "smallest" set of changes that will improve an individual’s outcome. The resulting algorithm is shown to be applicable to both supervised classification and counterfactual decision-making systems.