Learning Counterfactual Densities via Marginal Contrastive Discrimination

Aminata Ndiaye

We propose a method to estimate counterfactual densities in order to characterise causal effects beyond the average treatment effect (ATE). Interventional densities provide a richer description of outcome distributions but are challenging to estimate in the presence of high-dimensional confounding. To address this issue, we build on Marginal Contrastive Discrimination (MCD) [Meziani et al., 2026], which reframes conditional density estimation as a generalized contrastive learning task, enabling the use of supervised machine learning tools. This leads to a new framework that delivers accurate counterfactual density estimates. Through simulated numerical experiments, we demonstrate that the proposed approach effectively handles high-dimensional data and improves upon state-of-the-art methods, opening new perspectives for nuanced and robust causal analysis.