ID de Contribution: **133** Type: **Contributed talk**

It’s All in the Mix: Wasserstein Classification and Regression with Mixed Features

lundi 28 juillet 2025 16:30 (30 minutes)

Problem definition: A key challenge in supervised learning is data scarcity, which can cause prediction models to overfit to the training data and perform poorly out of sample. A contemporary approach to combat overfitting is offered by distributionally robust problem formulations that consider all data-generating distributions close to the empirical distribution derived from historical samples, where ‘closeness’ is determined by the Wasserstein distance. While such formulations show significant promise in prediction tasks where all input features are continuous, they scale exponentially when discrete features are present.

Methodology/results: We demonstrate that distributionally robust mixed-feature classification and regression problems can indeed be solved in polynomial time. Our proof relies on classical ellipsoid method-based solution schemes that do not scale well in practice. To overcome this limitation, we develop a practically efficient (yet, in the worst case, exponential time) cutting plane-based algorithm that admits a polynomial time separation oracle, despite the presence of exponentially many constraints. We compare our method against alternative techniques both theoretically and empirically on standard benchmark instances.
Managerial implications: Data-driven operations management problems often involve prediction models with discrete features. We develop and analyze distributionally robust prediction models that faithfully account for the presence of discrete features, and we demonstrate that our models can significantly outperform existing methods that are agnostic to the presence of discrete features, both theoretically and on standard benchmark instances.

Author: BELBASI, Mohammad Reza (Imperial College Business School)

Co-auteurs: M. SELVI, Aras (Imperial College Business School); M. WIESEMANN, Wolfram (Imperial College Business School)

Orateur: BELBASI, Mohammad Reza (Imperial College Business School)

Classification de Session: ML

Classification de thématique: Machine learning