

Mixed-feature Logistic Regression Robust to Distribution Shifts

mercredi 30 juillet 2025 10:45 (30 minutes)

Logistic regression models are widely used in the social and behavioral sciences and in high-stakes domains, due to their simplicity and interpretability properties. At the same time, such domains are permeated by distribution shifts, where the distribution generating the data changes between training and deployment. In this paper, we study a distributionally robust logistic regression problem that seeks the model that will perform best against adversarial realizations of the data distribution drawn from a suitably constructed Wasserstein ambiguity set. Our model and solution approach differ from prior work in that we can capture settings where the likelihood of distribution shifts can vary across features, significantly broadening the applicability of our model relative to the state-of-the-art. We propose a graph-based solution approach that can be integrated into off-the-shelf optimization solvers. We evaluate the performance of our model and algorithms on numerous publicly available datasets. Our solution achieves a 408x speed-up relative to the state-of-the-art. Additionally, compared to the state-of-the-art, our model reduces average calibration error by up to 36.19% and worst-case calibration error by up to 41.70%, while increasing the average area under the ROC curve (AUC) by up to 18.02% and worst-case AUC by up to 48.37%.

Authors: M. SUN, Qingshi; M. JUSTIN, Nathan; Prof. GOMEZ, Andres; Prof. VAYANOS, Phebe (University of Southern California)

Orateur: Prof. VAYANOS, Phebe (University of Southern California)

Classification de Session: ML

Classification de thématique: Machine learning