

Distribution shift inference

Advisor: Jan Miłniczuk

The research proposal concerns distribution shift i.e. discrepancy between distributions of training and testing population in classification and regression context. There are two special cases, label shift and covariate shift, which are intensively studied, the general case is less frequently considered due to methodological and technical issues. The main approach in a general case is based on density ratio estimation, which becomes ineffective when dimensionality of the space of predictors is high. The recent methods use dynamic weighting approach. The research will start from analysis of this method and investigate alternative approaches based on distribution matching, optimal transport in classical scenario, when both predictors and response are observable as well as in partial observability scenario such as Positive Unlabeled learning or corrupted data. The aim is to develop methodology, solutions and algorithms together with their efficient implementations.

A candidate is expected to have M.Sc. in Mathematics, Computer Science or Engineering, be knowledgeable in Machine Learning and Statistics, including both its mathematical and computational aspects, and possess sufficient computing skills to effectively implement and analyze proposed methods. Scientific curiosity and eagerness to learn are essential.

A prospective candidate should contact the author of the proposal before formal submission of documents (jan.mielniczuk@ipipan.waw.pl).

References

Fang, T. et al (2020) Rethinking of importance weighting for deep learning under distribution shift, NeurIPS 2020

Kumagai et al (2025) Importance weighted Positive-Unlabeled learning under distribution shift adaptation, AISTATS 2025

Maity, S. et al (2023) Understanding new tasks through the lens of training data via exponential tilting, ICLR 2023