Title: Going beyond the fear of emptyness to gain consistency

Missing data are ubiquitous in many real-world datasets as they naturally arise from gathering information from various sources in different format. Most statistical analyses have focused on estimation in parametric models despite missing values. However, accurate estimation is not sufficient to make predictions on a test set that contains missing data: a manner to handle missing entries must be designed. In this talk, we will analyze two different approaches to predict in presence of missing data: imputation and pattern-by-pattern strategies. We will show the consistency of such approaches and study their performances in the context of linear models.

Related papers:
- On the consistency of supervised learning with missing values https://arxiv.org/abs/2405.09196,
- What is a good imputation to predict with missing values https://arxiv.org/abs/2106.00311
- Near-optimal rate of consistency for linear models with missing values https://proceedings.mlr.press/v162/ayme22a/ayme22a.pdf
- Naive imputation implicitly regularizes high-dimensional linear models https://arxiv.org/abs/2301.13585
- Harnessing pattern-by-pattern linear classifiers for prediction with missing data https://arxiv.org/pdf/2405.09196