Causal Inference in Observational Data

Supervisor: Igor Kulev
igor.kulev@epfl.ch

July 13, 2018

1 Overview and Goal

Observational studies are rising in importance due to the widespread accumulation of data in fields such as healthcare, education, employment and ecology [1]. The information contained in these studies could be valuable to understand how we can tailor some interventions (actions designed to bring about a change in a process or an individual) to particular people in order to improve their health. This problem is challenging because the observational data is biased i.e. the actions observed in the data depend on variables which might also affect the outcome, resulting in confounding [2]. For example, richer patients might better affort certain medications. Recently, few methods based on neural networks were developed [1, 2, 3] that are able to extract causal relationships from observational data by first transforming the features into a new space in which the bias is removed, and then performing a regression on the transformed features.

The aim of this project is to implement and analyze one of the existing methods for causal inference in observational data, to test it on some benchmark and real-world datasets, and propose ways to improve its performance.

2 Required Skills

- Good knowledge of Machine Learning
- Experienced in Python

References

[1] O. Atan, J. Jordan, and M. van der Schaar, "Deep-treat: Learning optimal personalized treatments from observational data using neural networks." AAAI, 2018.

[2] U. Shalit, F. D. Johansson, and D. Sontag, "Estimating individual treatment effect: generalization bounds and algorithms," *arXiv preprint arXiv:1606.03976*, 2016.

[3] F. Johansson, U. Shalit, and D. Sontag, "Learning representations for counterfactual inference," in *International Conference on Machine Learning*, 2016, pp. 3020–3029.