EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATHEMATISCH-
NATURWISSENSCHAFTLICHE
FAKULTÄT

DATA SCIENCE & ANALYTICS RESEARCH GROUP
(DR. GJERGJI KASNECI)

Master Thesis EXPLAINABLE ML METHODS

Counterfactual Explanations in Causal ML Models

About the topic

As Machine Learning (ML) techniques grow more and more popular, they have started to support and execute decisions that have primarily been done by humans. For example, learning algorithms are being used by banks to evaluate how likely individuals are to pay back their loans. Likewise, in the US legal system As the influence of these risk assessments increase, decision makers would like both to **understand how complex ML algorithms make decisions** and to enable those with negative predictive outcomes to **change their score** and receive a positive outcome in the future.

Your task

Based on our previous work (see our website), you will use **Pearl's causal model framework** to study how inference on counterfactual explanations can be achieved under in the presence of **hidden confounders** and **model specification**. You will closely collaborate with Martin Pawelczyk and other students – currently working on related projects – to enhance our understanding of causal counterfactual **explanation models**. This thesis attempts to make first steps to close this gap in the literature and ideally results in a (NeurIPS, ICLR or ICML) workshop paper.

The main part of the thesis is to develop a transparent prediction framework. This includes:

- an experimental and theoretical understanding of counterfactual explanations in the presence of hidden confounders.
- an implementation of a new counterfactual explanation algorithm (provided by us) and its experimental evaluation

Requirements

Ideally you are motivated to work on **explainable ML methods**. Moreover, you ideally, but, not necessarily,

- have strong programming skills in Python;
- Have a sound background in machine learning (and statistics).

Contact

If you are interested, do not hesitate to approach us:

Martin Pawelczyk
Sand 14, C216
martin.pawelczyk(at)uni-tuebingen.de

Dr. Gjergji Kasneci
Sand 14, C221
gjergji.kasneci(at)uni-tuebingen.de

- November 2020