EBERHARD KARLS UNIVERSITÄT TÜBINGEN

DSAR
DATA SCIENCE & ANALYTICS RESEARCH

Master's thesis German or English

Generating Counterfactual Explanations for Image Data

Your task

In our group, we are interested in novel algorithms to explain the results of neural networks. Counterfactual Explanations constitute a popular explanation technique, especially for tabular data. The goal of this project is to transfer this approach to image data. This requires powerful generative models, e.g., Generative Adversarial Networks (GANs), diffusion models.
Thus, you will work on the following tasks:

- Literature research on existing approaches for the counterfactual explanation of image data
- Definition of required properties for practical counterfactual examples in the image domain
- Implementation of an algorithm that yields counterfactuals that fulfill the above properties
- Evaluation of the approach on real-world single- or multi-class datasets

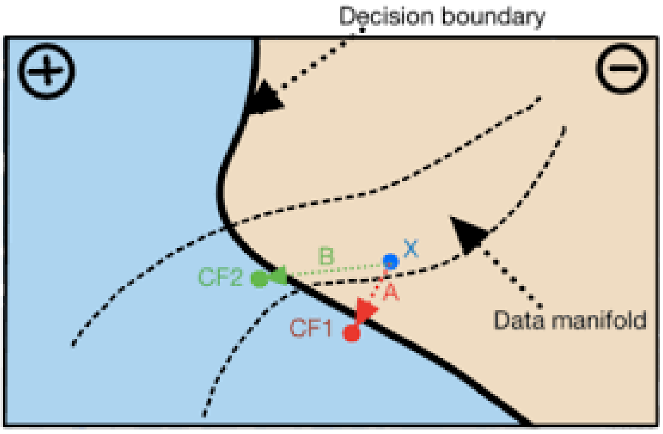
Your profile

- Background in computer science, mathematics of other related engineering majors
- Profound knowledge in Python programming and of machine learning algorithms

It is beneficial, though not mandatory, to have visited the "Data Mining & Probabilistic Reasoning" lecture or "Explainable & Fair Machine Learning" seminar by Prof. Dr. Gjergji Kasneci.

An ideal candidate has

- Practical experience working with GANs



- Strong interest in Machine Learning

We offer

- Intensive mentoring
- Opportunity to publish promising results in a research paper with us

After the successful completion of the thesis, you will be able to understand the state-of-the-art approaches for explaining the results of artificial neural networks and will know how to implement them using modern machine learning frameworks.

Are you interested?

Please send a short CV and current transcript of records to:

Tobias Leemann
Sand 14, C216
tobias.leemann@uni-tuebingen.de

Vadim Borisov
Sand 14, C207
vadim.borisov@uni-tuebingen.de

Prof. Dr. Gjergji Kasneci
Sand 14, C221
gjergji.kasneci@uni-tuebingen.de