Mathematical Underpinnings of Machine Learning - Project

Patryk Świątek

May 2024

1 Project Progress

- Selected topic: Feature selection (Project A).

- One-person team.

- I have familiarized myself with the literature addressed in the project description. I plan to implement and experiment with the information theory measures proposed in the literature, namely JMI and CMIM.

- I plan to use the arbitrary algorithms, such as:

    - Feature importances extracted from decision tree-based algorithms (XGBoost or Random Forest model).
    - Boruta algorithm.
    - Feature importance based on the SHAP (SHapley Additive exPlanations) values.

- For the experiments with the real data, I plan to use:

    - Statlog (German Credit Data) - available in Link.
    - Dataset with a large number of variables, e.g. genomic data.
    - Other dataset I will find interesting to this type of problem.