Project A - Feature Selection

Mathematical Underpinnings of Machine Learning

Bartosz Sawicki

May 2024

Checkpoint 1 - Current Status

I read about the mutual information-based feature selection and planned the following steps:

1. Feature selection methods to be implemented:
   - Joint Mutual Information (JMI)
   - Conditional Mutual Information Maximisation (CMIM)
2. Feature selection methods just to be used:
   - Regression with LASSO regularization
   - Recursive Feature Elimination using SVM as a base model
   - Fast-CMIM implementation from `https://github.com/CharlesGaydon/Fast-CMIM`
3. Real-world datasets:
   - Toxicity dataset (feature selection task) - `https://archive.ics.uci.edu/dataset/728/toxicity-2`
   - Period Changer (feature selection task) - `https://archive.ics.uci.edu/dataset/729/period+changer-2`
   - TUANDROMD malware detection dataset (only binary features) - `https://archive.ics.uci.edu/dataset/855/tuandromd+(tezpur+university+android+malware+dataset)`