Mathematical Underpinnings of Machine Learning
Project A - Checkpoint

Grzegorz Zakrzewski, 313555

The selected project is Project A - feature selection, which I am working on alone.

1 Progress

I have familiarized myself with the literature and have read Brown et al. [1]. After reviewing it, I have chosen the *Joint Mutual Information* (JMI) and *Conditional Mutual Info Maximisation* (CMIM) criteria. Additionally, I have explored the works of Yang and Moody [3] and Fleuret [2] to deepen my understanding.

I have implemented the JMI and CMIM criteria according to the formulas provided by Brown et al. The implementation was done in Python and relies on the `mutual_info_score` function from the `scikit-learn` library. Additionally, I have devised a stopping rule, which aborts the feature selection search when a large percentage of the total number of features is selected or when the value of the JMI or CMIM criteria drops rapidly.

As for arbitrary methods, I have prepared one method based on Random Forest feature importance. However, I still need to come up with a second method.

I have created four artificial datasets based on random samples from a normal distribution. The first contains 10 relevant and 10 completely irrelevant features. In the second dataset, every feature is duplicated, resulting in ten pairs of identical features. In the third dataset, there are also 10 pairs of features, but they are not identical; rather, they are highly correlated, and only one feature of each pair is relevant. In the last dataset, 5 out of 10 relevant variables have additional Gaussian noise added. I still need to prepare an artificial example in which one of the MI-based methods doesn't work.

I am currently planning how to conduct and assess experiments to ensure a fair comparison.

References

[1] Gavin Brown, Adam Pocock, Ming-Jie Zhao, and Mikel Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *The journal of machine learning research*, 13(1):27–66, 2012.

[2] François Fleuret. Fast binary feature selection with conditional mutual information. *Journal of Machine learning research*, 5(9), 2004.

[3] H Yang and John Moody. Feature selection based on joint mutual information. In *Proceedings of international ICSC symposium on advances in intelligent data analysis*, volume 23. Citeseer, 1999.