Assignment 1
The Scientific Discourse

Tymur Mykhalievskyi
7031100

April 21, 2025

1 General Overview

2 MIC

In this section we will share my thoughts on MIC. The overall idea seems promising on the first glance. MIC is supposed to be the one metric that is general and equitable (according to the definition by Reshef et al. (2011)). Firstly reading the work by Reshef et al. (2011), prima facie, MIC completely satisfies these properties. In particular, MIC does not make any assumptions about the underlying funciton which is supposed to make it generalize very well. However, after reading the concerns presented by Simon and Tibshirani (2014) and Kinney and Atwal (2014), a more thorough analysis of MIC had to be conducted.

2.1 Limitations of MIC

2.1.1 Grid Resolution

2.1.2 Geometrical Interpretation of MIC

The underlying implementation of MIC is based on the idea of partitioning the scatter plot of two variables into a grid. At this point we can see an important limitation of MIC which also does not align with the equitability of the metric. Due to the fact that the grid is made of rectangles, the relationship some relationships between two variables could work particularly well. For example, if the underlying function is a step function, the grid would capture the relationship very well. However, if we exchange the underlying function to a modular function, the grid would not be able to capture the realtionship as well. This is because the step function would be able to fit into the grid cells very well, while the modular function being not parallel to any of the axis would not be able to fit into the grid cells. This parituclar case is visualized in Figure 1

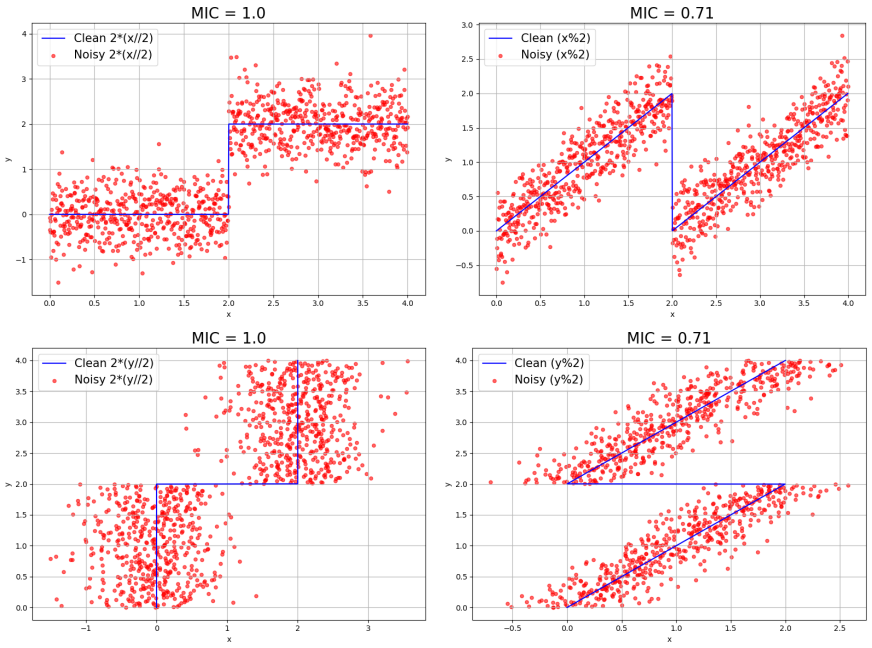

Figure 1: Comparison of the step function and modular function. The left plot shows the step function with its inverse variant on the bottom. While the right plot shows the modular function with its inverse variant. Both plots were generated using the exact same parameters: $n = 1000, 1 - R^2 = 0.2, x \in [0, 4)$. The source code can be accessed via GitHub (Mykhalievskyi, 2025)

2.2 Why MIC will not be the best

3 Equitability

References

Kinney, J. B., & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, *111*(9), 3354-3359. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1309933111` doi: 10.1073/pnas.1309933111

Mykhalievskyi, T. (2025). *Source code for the plots.*

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., ... Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science*, *334*(6062), 1518-1524. Retrieved from `https://www.science.org/doi/abs/10.1126/science.1205438` doi: 10.1126/science.1205438

Simon, N., & Tibshirani, R. (2014). *Comment on "detecting novel associations in large data sets" by reshef et al, science dec 16, 2011.* Retrieved from `https://arxiv.org/abs/1401.7645`