# CMI estimation
# project B

Julia Girtler
Helena Wałachowska

Wydział Matematyki i Nauk Informacyjnych PW
SMAD

Warszawa, 28 stycznia 2025

# Presentation plan

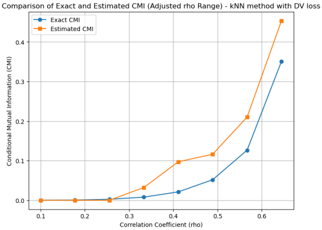
# Three estimators

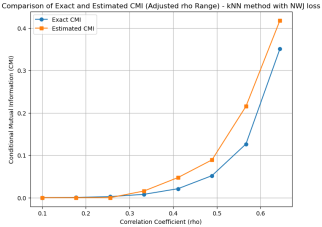
1. kNN methon by J. Runge
   - kNN method for creating sample $X_{perm}$, so that the joint distributions $(X_{perm}, Y)$ and $(X_{perm}, Z)$ remain equal to the distributions $(X, Y)$ and $(X, Z)$, respectively. (we used 5 NN)
   - neural network for CMI estimation using KL divergence (two variants Donsker-Varadhan or the Nguyen-Wainwright-Jordan loss function). We estimate the distance between $(X, Y, Z)$ (from the distribution $p(x, y, z)$) and $(X_{perm}, Y, Z)$ (from the distribution $p(x|z)p(y|z)p(z)$).
2. chain rule
   - neural network for estimating the value of $I((X, Z)|Y)$ as the KL divergence between the distributions of the samples $(X, Z, Y)$ and $(X, Z, Y_{perm})$, and the value of $I((X)|Y)$ as the KL divergence between the distributions of $(X, Y)$ and $(X, Y_{perm})$.
   - We calculate CMI using formula $I(X, Y|Z) = I((X, Z)|Y) - I((X)|Y)$.
3. MI estimator from sklearn
   - mutual_info_regression(X, Y, discrete_features=False, n_neighbors=5) function from the sklearn package

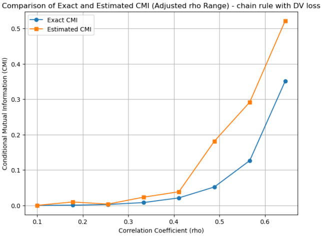
# Data

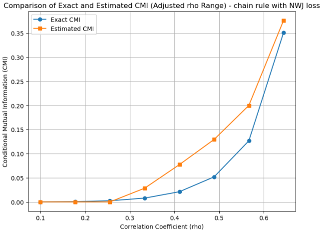
1. Data from normal distribution, for which the formula for CMI is known
2. The dataset that consists of 10000 samples and 20 features, generated using the `make_regression` function, features have been manually manipulated to have strong relationships with others
3. The California Housing dataset that consists of various features with the target variable being the median house value (Y)

# Results of data 1



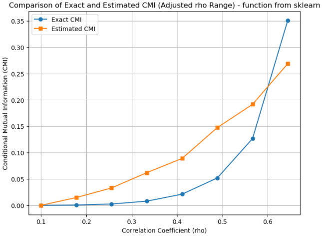



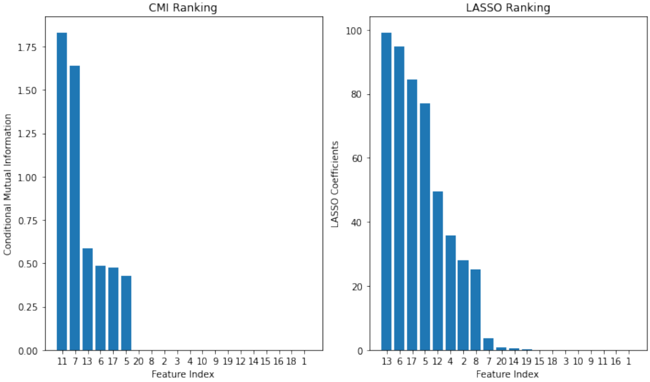



# Results of data 1

- We verified that all proposed estimators perform well in detecting the absence of conditional dependence in the data.
- We observe that both Estimator 1 and 2 perform well in this scenario. In both cases, when using the DV loss function and the NWJ method, there is some instability in the results.

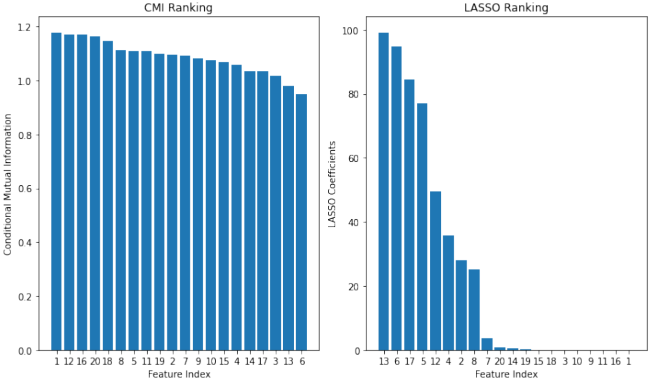


# Results of data 2 - Estimator 1



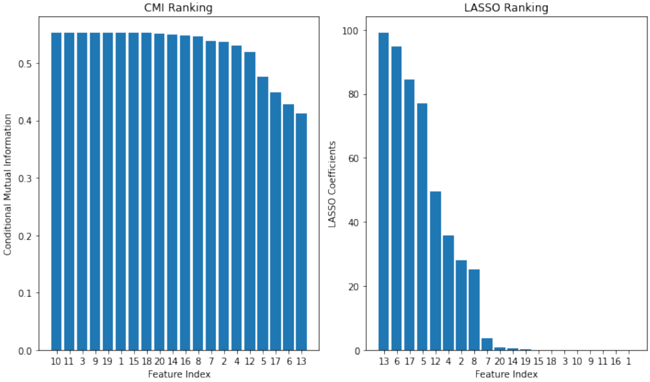

**Number of inversions:** 78
**Top 10 agreement score:** 0.8

# Results of data 2 - Estimator 2




**Number of inversions:** 95
**Top 10 agreement score:** 0.5

# Results of data 2 - Estimator 3




**Number of inversions:** 106
**Top 10 agreement score:** 0.1

# Main conclusions

- In the first experiment, it was verified that all proposed estimators perform well in detecting the absence of conditional dependence in the data
- In the case of artificially generated data with the problem of selecting the 10 most significant features, Estimator 1 performed the best, achieving 70-80% agreement with the LASSO method. Estimator 2 performed slightly worse, achieving 40-50% agreement. Estimator 3 made a decisively different selection, achieving only 10-20% agreement with LASSO.