Project B **Mathematical Underpinnings**

CMI estimation

Methods

The goal of this project is to implement estimators of conditional mutual information (CMI) and use them to rank features based on their importance. We will compare the estimation of CMI for continuous random variables using the following methods:

- **generative and divergence based CMI estimation** (your own implementation)
  - first step is using an algorithm estimating conditional distribution $p(y|z)$ and thus also factorised joint distribution under conditional independence $p(x|z)p(y|z)p(z)$ (e.g. kNN based permutation:Runge; other methods can be found e.g. in Mukherjee et al),
  - next step is neural estimation of KL divergence based on a variational representation of KL divergence of your choice e.g. Donsker-Varadhan.
- **difference based CMI estimation** (the implementation can be based on code from laboratories)
  - first, use chain rule $I(X;Y|Z) = I(X,Z;Y) - I(X;Y)$.
  - next, estimate two MIs on the right hand side using a neural estimator of KL divergence based on a variational representation of KL divergence of your choice e.g. Donsker-Varadhan.
- any other estimator (might be already implemented and ready to use)

Using more than one variant of an estimator in each bullet point is highly encouraged, although not obligatory.

Evaluation of the methods

At least one example in each category must be provided:

1. an example, for which the formula for CMI is known and exact CMI can be computed (on a grid of parameters),
2. a simulation example of ranking variable importance based on your estimators,
3. a dataset example of ranking variable importance.

Ad 2 and 3: Consider 10 most important variables.
Ad 2 and 3: Devise your own measure of quality of ranking based e.g. on the number of inversions or the number of cases when the most important k variables have been chosen.