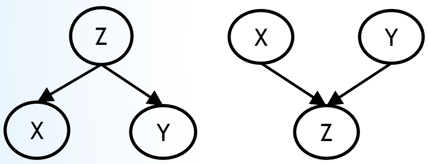
Tell Causality from Spurious Correlation Induced by Confounding

Introduction: Causal relationship is different from statistical relationship. Variables statistically related may be causally unrelated. Independence test can be used to test whether two variables are statistically independent. However, it cannot distinguish spurious correlation caused by confounding. There are two confounding cases : common cause and selection bias. We proposed an original way to tell the two confounding cases from real causality.

Confounding:



(1) Common Cause (2) Selection Bias

Fig. Two confounding cases where Z may induce spurious correlation between X and Y

Proposed Method:

- Infer intrinsic dimension of data[1]

$$C(\epsilon) = \lim_{N\to\infty} \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H(\epsilon - \|x_i - x_j\|)$$

$$H(u) = \begin{cases} 0, \; if \; u < 0 \\ 1, \; if \; u \geq 0 \end{cases}$$

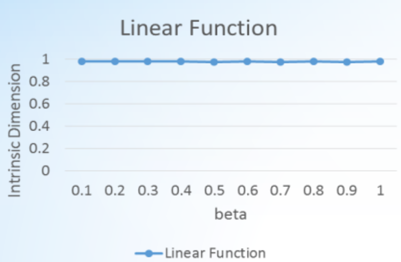
$$d = \lim_{\epsilon\to 0} \frac{logC(\epsilon)}{log\epsilon} \approx \frac{\log(C(\epsilon_1) - C(\epsilon_2))}{\log(\epsilon_1 - \epsilon_2)}$$

- Decide whether Z is confounder according to the inferred result.

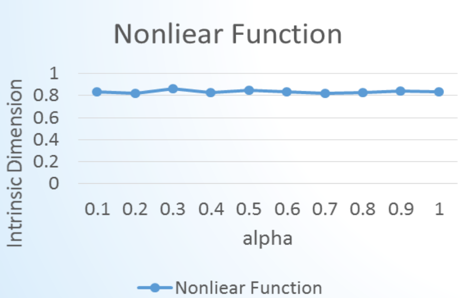
Simulation:

Common Cause Case:

(1) Linear Function



(2) Non-linear Function



Real World Data:

- Altitude, temperature and precipitation

This is an example of common cause case where altitude is the root variable Z. The inferred intrinsic dimension is 1.0534, which is close to the ideal case.

- Cement, Blast Furnace Slag, Compressive Strength

This is an example of selection bias case in Fig. 1 where compressive strength is the common effect Z. The inferred intrinsic dimension is 1.9640.

Conclusion and future work:

We proposed to use intrinsic dimension estimation to distinguish confounding from real causality heuristically. However, we found some limitations of the existing method through experiments, our future work is to improve the method.

Reference:

1. Eckmann, J-P., and David Ruelle. "Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems." *Physica D: Nonlinear Phenomena* 56.2-3 (1992): 185-187.