Canonical Correlation Analysis for Causal Inference via Generative Modeling

IA pour la Science

December 27^{th}, 2025

The code of the computational experiment

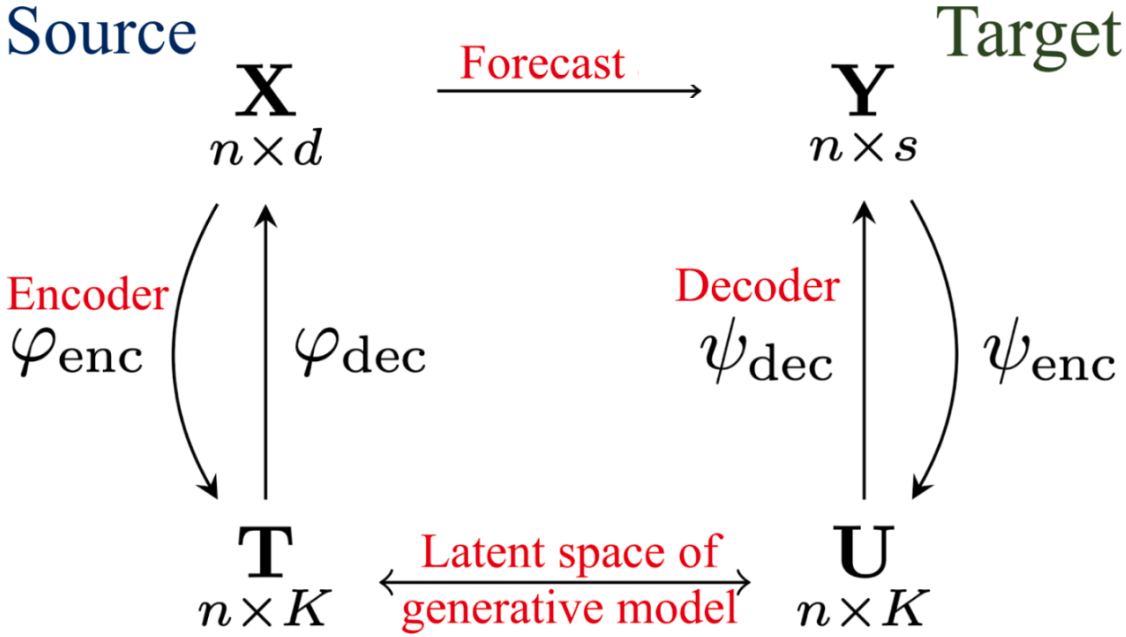
List of the modules for minimum working example

1. The model for seq2seq forecasting: 1) CCA, 2) transformer, 3) neural ODE
2. The algorithm to generate intervention:
   1) do-calculus deterministic strategy,
   2) kNN,
   3) Convergent Cross-Mapping
   4) diffusion
3. The criterion to assess causality:
   1) statistical significance of the distribution change,
   2) KL-divergence

The datasets

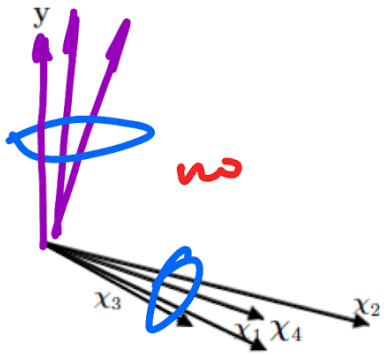
1. Observed multivariate time series d-variate to r-variate $d > r$
2. Text data represented as multivariate time series

Generative model in canonical correlation analysis



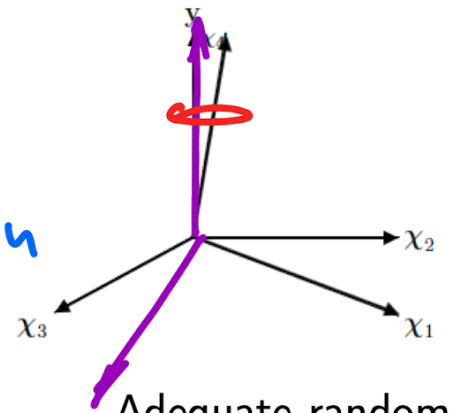
The source and the target reconstructs the observable variables. Since the both variables generated by a common cause they share the same latent space. The idea is to infer the causal relationship through the change of the distribution of the latent variables under the intervention.

Configuration of self-reconstruction and causality

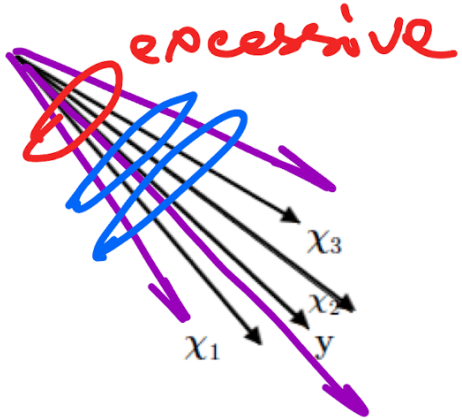


Non-adequate correlated

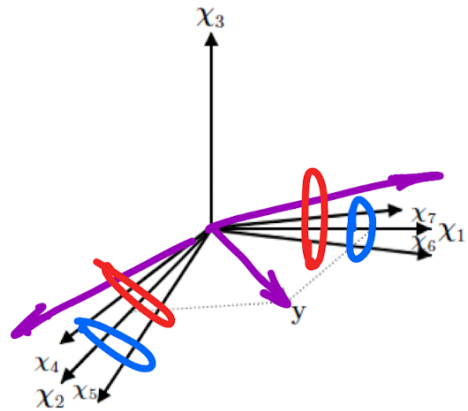
y target
y source
• self-reconstruction
• causality



Adequate random



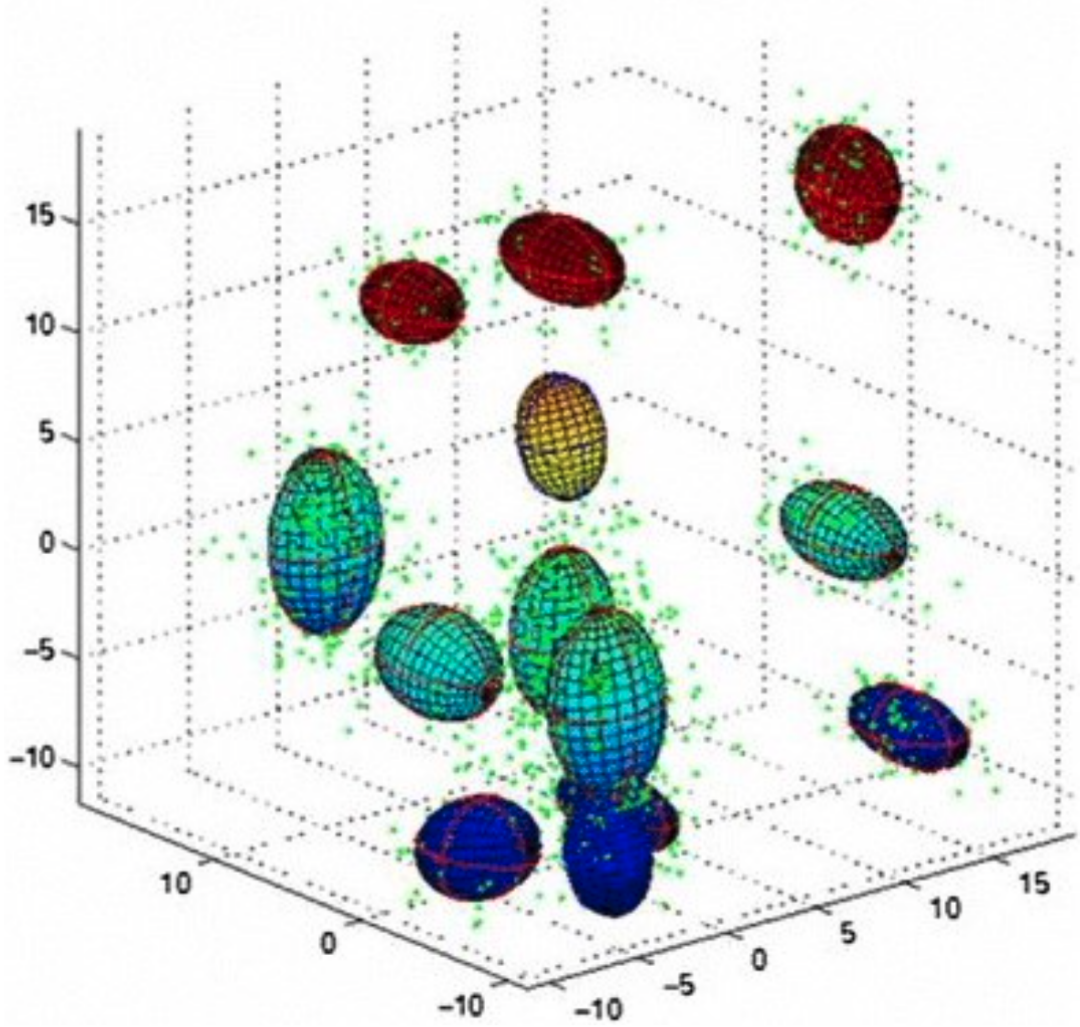
Adequate redundant



Adequate correlated

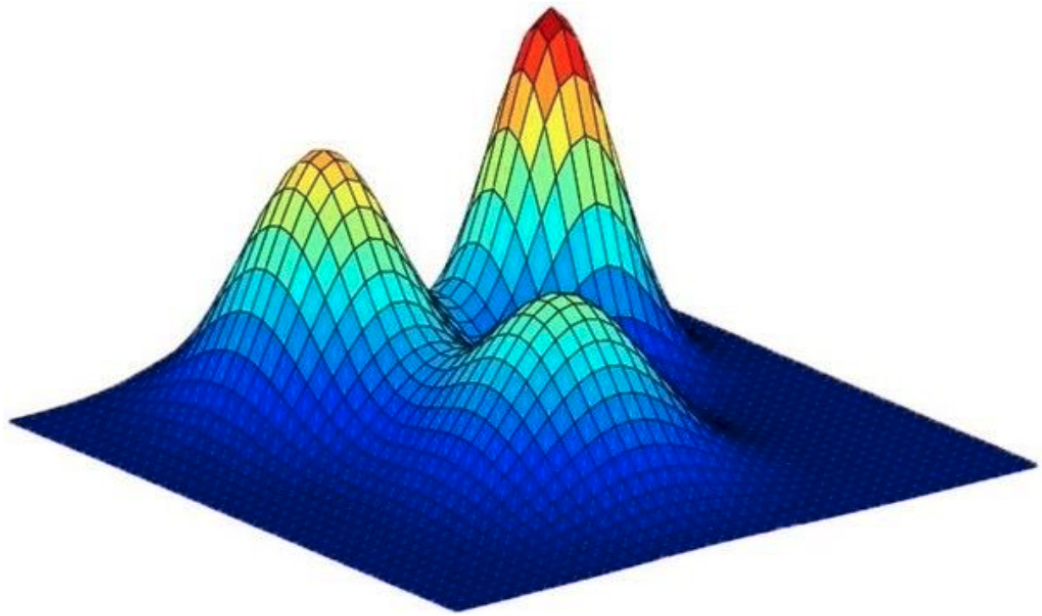
Find the causal relationship in the set of the reconstructed components in the source and the target. Assume the simplest causal DAG is bipartite. It connects some subset of the source components to some subset of the target components.

Disribution of encoder components



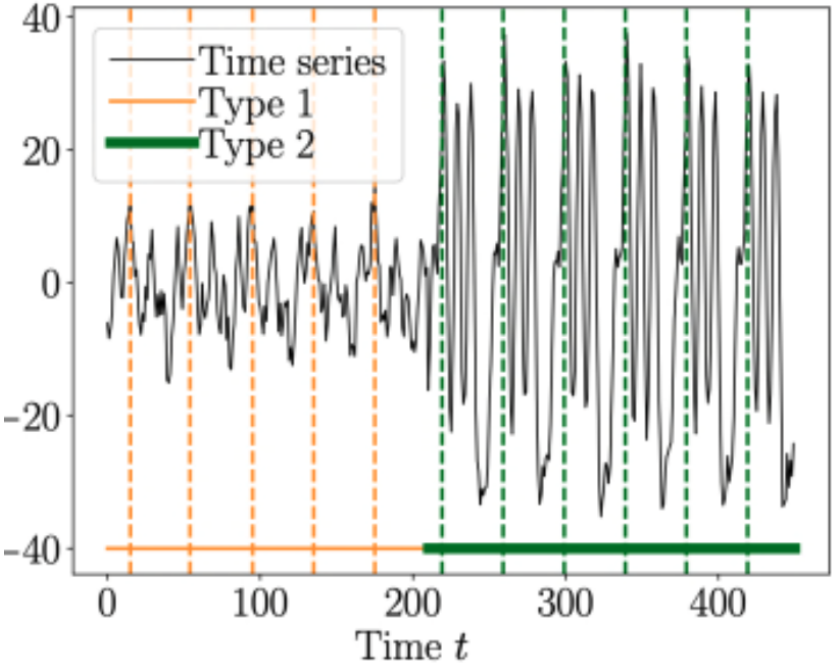
Each neuron w_j from the autoencoder with parameters $[\dots w_j^T \dots]^T = W^T$ sits in the same probabilistic space $w_j \sim \mathcal{N}(\hat{w}_j, A_j)$ due to the dot product $W^T x$.

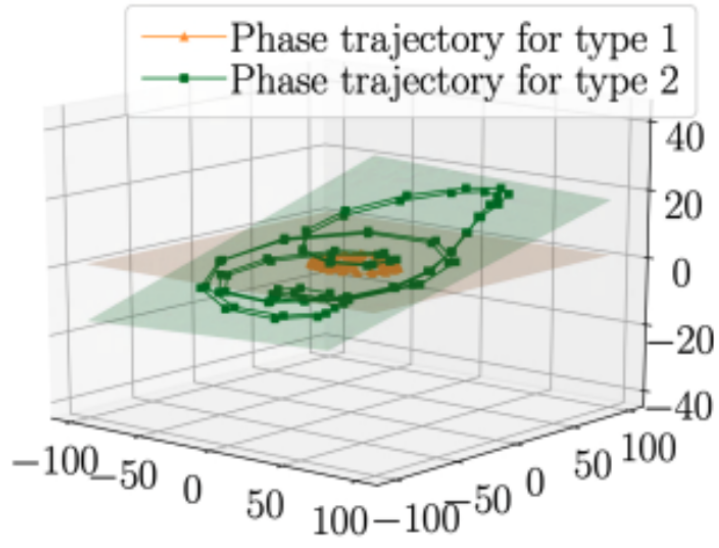
Mixiture of distributions for series of interventions

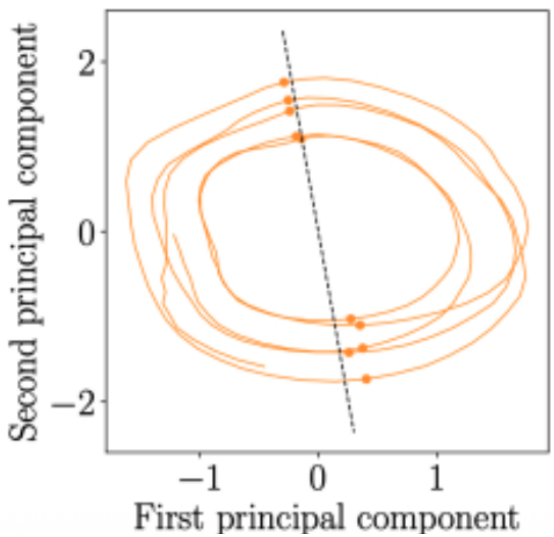


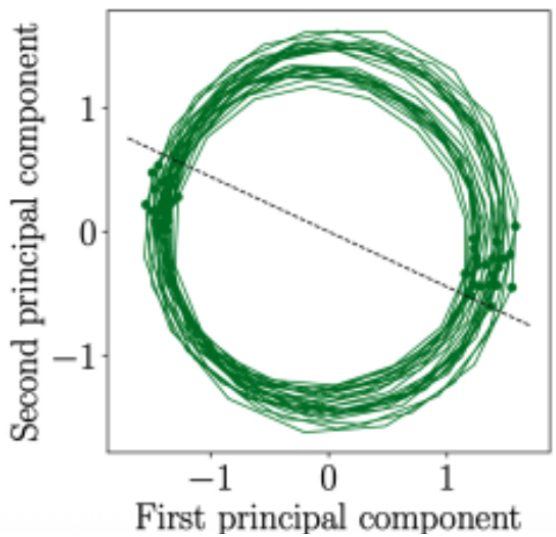
Statistical significance of the change in the distribution of the latent variables under the intervention indicates the presence of a causal relationship between the time series.

Change of the phase trajectory to make interventions









The intervention is performed by changing the phase trajectory of the source time series.

Questions to discuss

1. Configurations of reconstruction and causal models
2. Generative model as a form of intervention
3. List of models for decoding of time series and text data