Computational Hypergraph Discovery

Théo Bourdais

July 17, 2025

California Institute of Technology

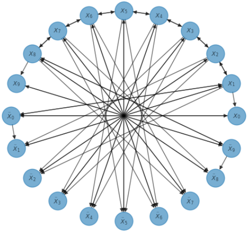
x_1	$\ddot{x}_1$	...	$\ddot{x}_{10}$
0.45	0.66	...	-0.23
⋮	⋮	⋱	⋮
-0.78	-0.12	...	0.89

$$\xrightarrow{\text{CHD}}$$

$$\ddot{x}_1 = \frac{c^2}{h^2}(x_2 + x_0 - 2x_1)(1 + (x_2 - x_0)^3)$$

$$\vdots$$

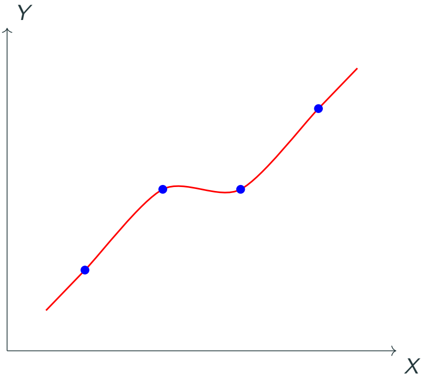
$$\ddot{x}_{10} = \frac{c^2}{h^2}(x_9 - 2x_{10})(1 - x_9^3)$$



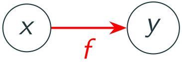
Codiscovering graphical structure and functional relationships within data: A Gaussian Process framework for connecting the dots. Bourdais, T., Batlle, P., Yang, X., Baptista, R., Rouquette, N., Owhadi, H. (2024).
Proceedings of the National Academy of Sciences

The regression problem

Suppose $y = f(x)$, given samples (X_i, Y_i) for $i = 1, .., N$, approximate f



Graph representation



For $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, .., N$, approximate f s.t. $y = f(x)$.

Linear Ridge regression

Our approximation is a linear function $\tilde{f}(x) = \beta^{*T} x$ with

$$\beta^* = \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{N} |Y_i - \beta^T X_i|^2 + \gamma \|\beta\|^2$$

We know that, for $k(x, y) = x^T y$,

$$\tilde{f} \in \mathcal{H}_k = \{\text{Linear functions}\} = \{\sum_i \alpha_i k(\cdot, z_i), \text{for some } z_i, \alpha_i\}$$

For $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, .., N$, approximate f s.t. $y = f(x)$.

Quadratic Ridge regression

Our approximation is a quadratic function $\tilde{f}(x) = {\beta^*}^T \psi(x)$, $\psi(x) = (1, x, x^2)$,

$$\beta^* = \underset{\beta \in \mathbb{R}^{2p+1}}{\arg\min} \sum_{i=1}^{N} |Y_i - \beta^T \psi(X_i)|^2 + \gamma \|\beta\|^2$$

We know that, for $k(x, y) = \psi(x)^T \psi(y)$,

$$\tilde{f} \in \mathcal{H}_k = \{\text{Quadratic functions}\} = \{\sum_i \alpha_i k(\cdot, z_i), \text{for some } z_i, \alpha_i\}$$

Caltech

For $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$, $i = 1, .., N$, approximate f s.t. $y = f(x)$.

Kernel Ridge regression

Our approximation is a function in a space $\mathcal{H}_k$[1] defined by the kernel k.

$$\tilde{f} = \arg\min_{f \in \mathcal{H}_k} \sum_{i=1}^{N} |Y_i - f(X_i)|^2 + \gamma \|f\|^2$$

We know that,

$$\tilde{f} \in \mathcal{H}_k = \overline{\{\sum_i \alpha_i k(\cdot, z_i), \text{for some } z_i, \alpha_i\}}$$

[1] $\mathcal{H}_k$ is called the Reproducing Kernel Hilbert Space (RKHS) of k

Computational Hypergraphs

A computational hypergraph is a graphical representation of a set of equations

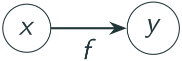
$$x \xrightarrow[f]{} y$$

$$y = f(x)$$

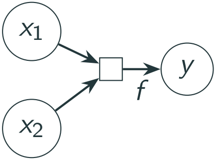
From Regression to Hypergraph discovery

Computational Hypergraphs

A computational hypergraph is a graphical representation of a set of equations



$$y = f(x)$$

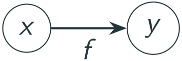


$$y = f(x_1, x_2)$$

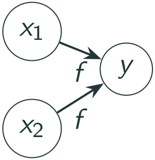
From Regression to Hypergraph discovery

Computational Hypergraphs

A computational hypergraph is a graphical representation of a set of equations



$$y = f(x)$$



$$y = f(x_1, x_2)$$

Computational Hypergraphs

A computational hypergraph is a graphical representation of a set of equations

$$\begin{tikzcd} x \arrow[r, "f"'] & y \end{tikzcd}$$

$$y = f(x)$$

$$\begin{tikzcd} x_1 \arrow[dr, "f"'] & \\ & y \\ x_2 \arrow[ur, "f"'] & \end{tikzcd}$$

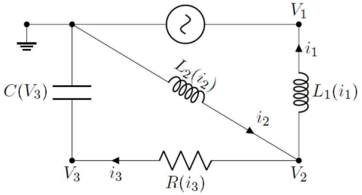
$$y = f(x_1, x_2)$$

$$\begin{tikzcd} x \arrow[r, "f"'] & y \arrow[r, "g"'] & z \end{tikzcd}$$

$$y = f(x)$$

$$z = g(y)$$

The electrical circuit example[2]



$$i_1 + i_3 = i_2$$

$$i_3 = C(V_3)\frac{dV_3}{dt}$$

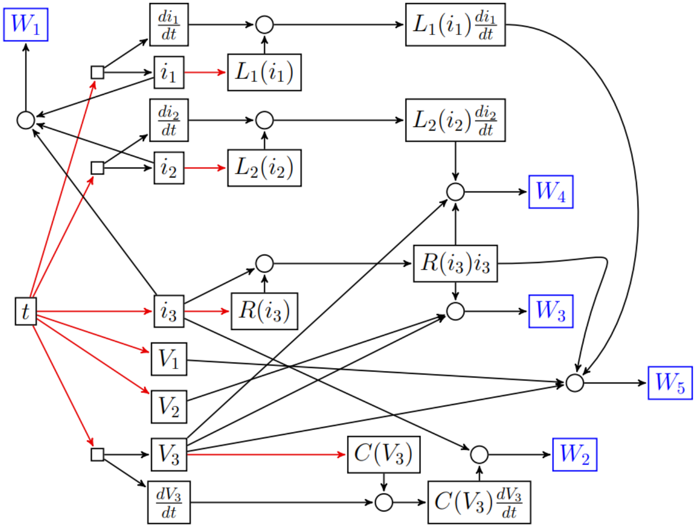
$$V_2 - V_3 = R(i_3)i_3$$

$$-V_2 = L_2(i_2)\frac{di_2}{dt}$$

$$V_2 - V_1 = L_1(i_1)\frac{di_1}{dt}$$

[2] Owhadi, *Computational Graph Completion*.

Since any set of equations can be represented as a Computational Hypergraph, we can obtain:

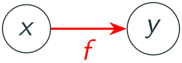


From Regression to Hypergraph discovery

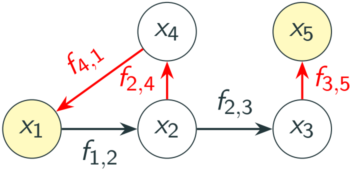
Regression Given samples $Y_i = f(X_i)$ for $i = 1, .., N$, approximate f

$$x \xrightarrow{f} y$$

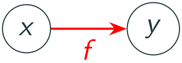
Regression Given samples $Y_i = f(X_i)$ for $i = 1, .., N$, approximate f



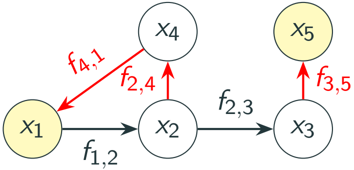
Hypergraph Completion Given the graph's structure and samples of its variables, approximate unknown edges, and missing data.



Regression Given samples $Y_i = f(X_i)$ for $i = 1, .., N$, approximate f



Hypergraph Completion Given the graph's structure and samples of its variables, approximate unknown edges, and missing data.



Hypergraph discovery Given samples of the variables, find the structure of the graph.

- Brain networks

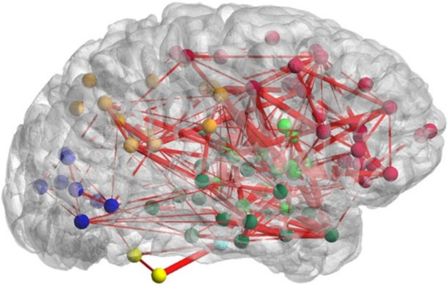


Figure 1: Image from Shu-Hsien Chu et al.

Objective: Discover functional dependencies between the activities of different brain regions.

- Brain networks
- Economic networks

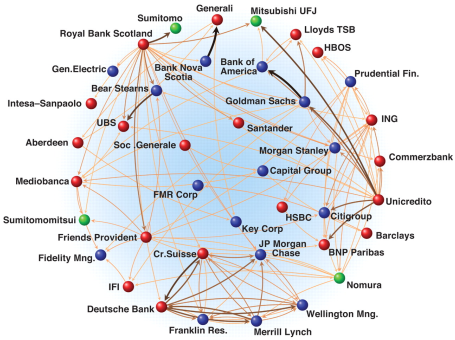


Figure 1: Image from Schweitzer et al.

Objective: Discover functional dependencies between economic markers of different banks

- Brain networks
- Economic networks
- Weather modelling

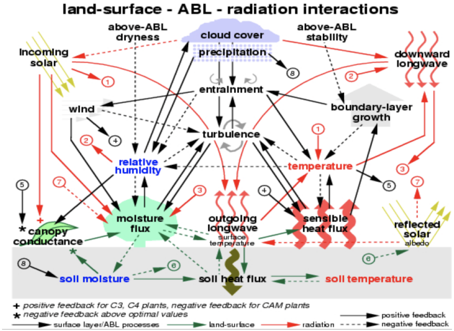


Figure 1: Image from Michael Ek.

Objective: Discover functional dependencies between the different variables

Causal inference and Probabilistic graphs

- Usually tackle a different problem (e.g., conditional independence or causality)
- Relies on strong assumptions, e.g.
  - Access to a distribution
  - Control on sampling
  - Ordering of variables (very hard to find)
- Computational complexity can blow up

Causal inference and Probabilistic graphs

- Usually tackle a different problem (e.g., conditional independence or causality)
- Relies on strong assumptions, e.g.
  - Access to a distribution
  - Control on sampling
  - Ordering of variables (very hard to find)
- Computational complexity can blow up

Sparse regressions

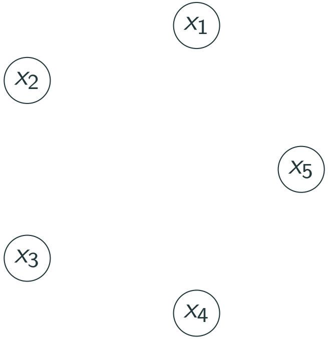
- Uses knowledge of sparse representations in a dictionary of functions
- Example: SINDY

Caltech

The CHD problem

Given N samples of our variables, recover the functional dependencies between them (i.e., the structure of the graph).

Starting from an empty graph

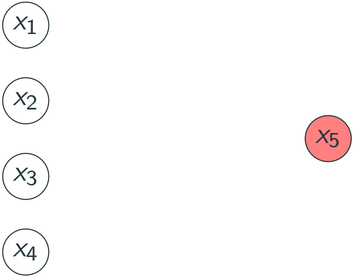


The CHD problem

Given N samples of our variables, recover the functional dependencies between them (i.e., the structure of the graph).

Ancestors: If $x_5 = f(x_1, x_4)$ for some f, x_1, x_4 are ancestors of x_5.

Ancestors: If $x_5 = f(x_1, x_4)$ for some f, x_1, x_4 are ancestors of x_5.

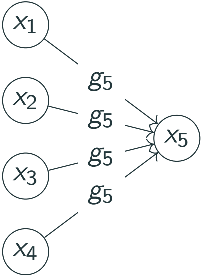


Our solution

x_5 is not a function of

$x_1, \ldots, x_4$

or



There is a function g_5 s.t.

$$x_5 = g_5(x_1, \ldots, x_4)$$

There are three questions:

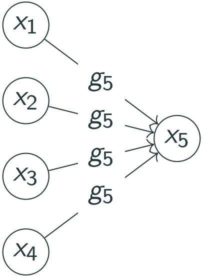
- Does x_5 have any ancestors?

Our solution

x_5 is not a function of

$x_1, \ldots, x_4$

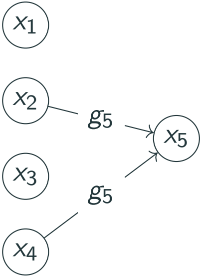
or



There is a function g_5 s.t.

$x_5 = g_5(x_1, \ldots, x_4)$

or



There is a function g_5 s.t.

$x_5 = g_5(x_2, x_4)$

There are three questions:

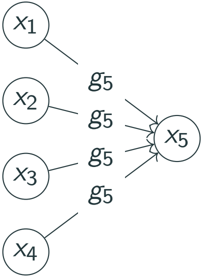
- Does x_5 have any ancestors?
- If so, what is the minimum set of ancestors?

Our solution

x_5 is not a function of

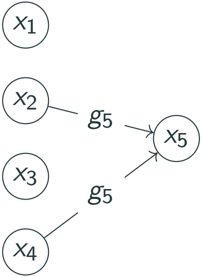
$x_1, \ldots, x_4$

or



There is a function g_5 s.t.

$x_5 = g_5(x_1, \ldots, x_4)$



There is a function g_5 s.t.

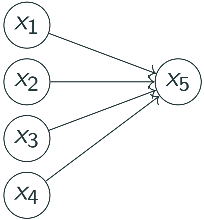
$x_5 = g_5(x_2, x_4)$

There are three questions:

- Does x_5 have any ancestors?
- If so, what is the minimum set of ancestors?
- What kind of function is g_5?

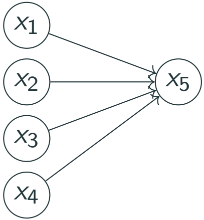
Does x_5 have any ancestors?

Ancestors,
samples gathered in X



Target
samples gathered in Y

Ancestors,
samples gathered in X



Target
samples gathered in Y

Let's see if there is g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a Gaussian Process (kernel k and noise variance γ):

$$g_5 = \arg\min_f \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2 \tag{1}$$

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \ \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s = \|g_5\|_k^2$$

(variance of data explained by model)

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s \;=\; \|g_5\|_k^2$$

(variance of data explained by model)

$$n \;=\; \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \; \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s \;=\; \|g_5\|_k^2$$

(variance of data explained by model)

$$n \;=\; \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

The noise n comes from two sources:

- True noise from the data

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \ \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s = \|g_5\|_k^2$$

(variance of data explained by model)

$$n = \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

The noise n comes from two sources:

- True noise from the data
- Unexplained data variance

Does x_5 have any ancestors?

Caltech

To see if this model correctly describes the data, we perform a nonlinear variance decomposition:

$$g_5 = \arg\min_f \ \|f\|_k^2 + \frac{1}{\gamma}|f(X) - Y|^2$$

$$s = \|g_5\|_k^2$$

(variance of data explained by model)

$$n = \frac{1}{\gamma}|g_5(X) - Y|^2$$

(variance of data explained by noise)

The noise n comes from two sources:

- True noise from the data
- Unexplained data variance
    - Quantifies model misspecification

Does x_5 have any ancestors?

Noise-to-signal ratio

$\frac{n}{n+s} \in [0,1]$, quantifies how much the data agrees with x_5 having $x_1, .., x_4$ as ancestors

- if $\frac{n}{n+s} \approx 0$: The model is well specified.
- if $\frac{n}{n+s} \approx 1$: The model is misspecified.

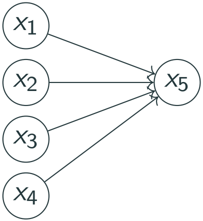
Does x_5 have any ancestors?

Noise-to-signal ratio

$\frac{n}{n+s} \in [0, 1]$, quantifies how much the data agrees with x_5 having $x_1, .., x_4$ as ancestors

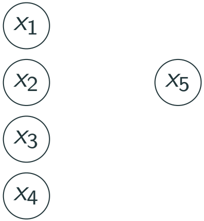
- if $\frac{n}{n+s} \approx 0$: The model is well specified.
- if $\frac{n}{n+s} \approx 1$: The model is misspecified.

If $\frac{n}{n+s} < 0.5$, x_5 has ancestors



or

If $\frac{n}{n+s} > 0.5$, x_5 has no ancestors

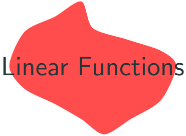


Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.

$\boxed{g_5 \text{ linear}}$



Current kernel: Linear

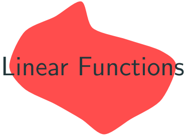
$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.

$$\boxed{g_5 \text{ linear}} \xrightarrow{\frac{n}{n+s} < 0.5}$$

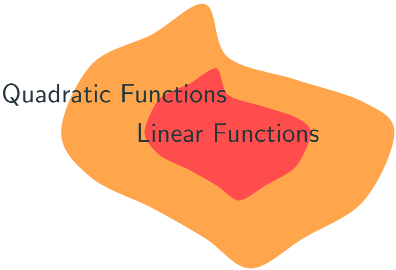
x_5 has ancestors and g_5 is linear

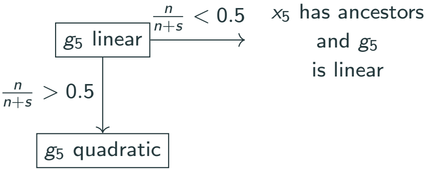


Current kernel: Linear

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i$$

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.



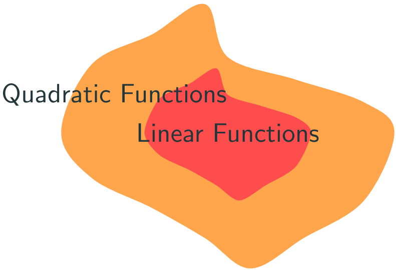


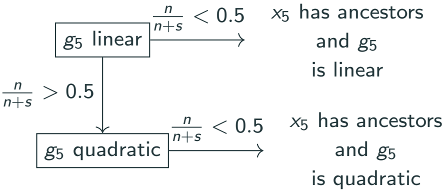
Current kernel: Quadratic

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.

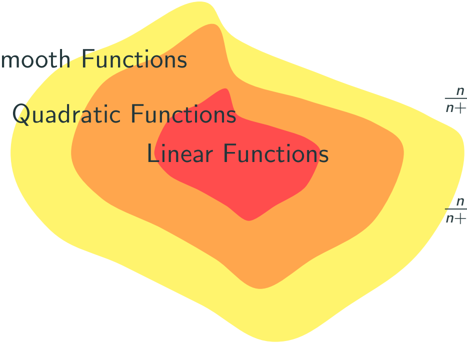


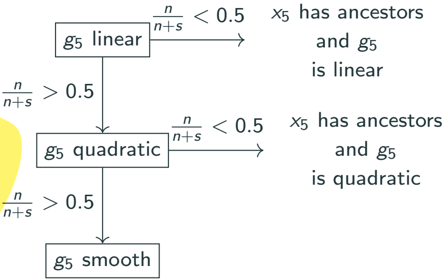


Current kernel: Quadratic

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.

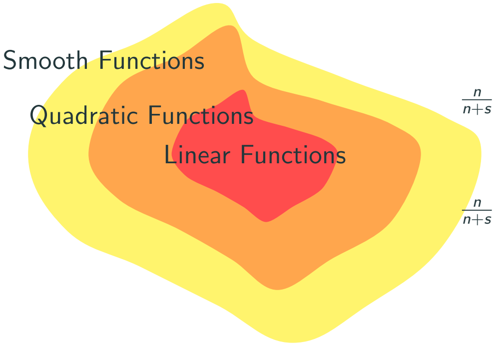




Current kernel: **Nonlinear**

$$k(x,y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j + \prod_{i=1}^{n} (1 + e^{-(x_i - y_i)^2})$$

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.



g_5 linear $\xrightarrow{\frac{n}{n+s} < 0.5}$ x_5 has ancestors and g_5 is linear

$\frac{n}{n+s} > 0.5$ ↓

g_5 quadratic $\xrightarrow{\frac{n}{n+s} < 0.5}$ x_5 has ancestors and g_5 is quadratic

$\frac{n}{n+s} > 0.5$ ↓

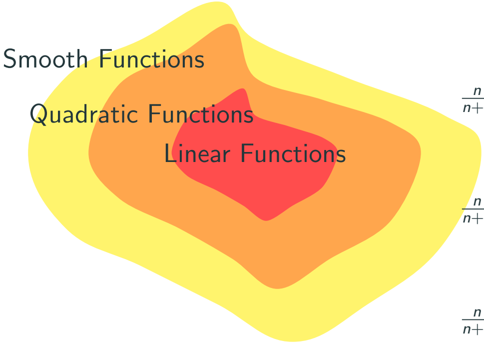
g_5 smooth $\xrightarrow{\frac{n}{n+s} < 0.5}$ x_5 has ancestors and g_5 is smooth

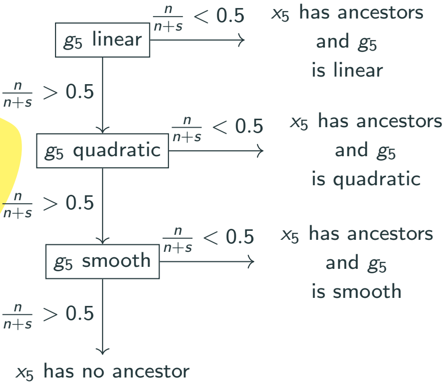
Current kernel: Nonlinear

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j + \prod_{i=1}^{n} (1 + e^{-(x_i - y_i)^2})$$

Choosing the kernel

Is there a g_5 s.t. $x_5 = g_5(x_1, .., x_4)$? The kernel defines the set of functions in which we are searching g_5.





Current kernel: **Nonlinear**

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j + \prod_{i=1}^{n} (1 + e^{-(x_i - y_i)^2})$$

To find the minimum set of ancestors, we will:

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:

what is the minimum set of ancestors?

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
  - Identify the least important potential ancestor

what is the minimum set of ancestors?

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
    - Identify the least important potential ancestor
    - Remove it from potential ancestors

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
  - Identify the least important potential ancestor
  - Remove it from potential ancestors
  - Check that the data agrees

Caltech

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
    - Identify the least important potential ancestor
    - Remove it from potential ancestors
    - Check that the data agrees $\Leftarrow$ we use $\frac{n}{n+s}$

To find the minimum set of ancestors, we will:

- Start with all nodes as potential ancestors
- Iteratively prune potential ancestors:
  - Identify the least important potential ancestor
  - Remove it from potential ancestors
  - Check that the data agrees $\Leftarrow$ we use $\frac{n}{n+s}$

Caltech

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Observe $k = k_2 + k_{-2}$

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Observe $k = k_2 + k_{-2}$

- k_2 depends on x_2: $k_2(x, y) = x_2 y_2 + x_2^2 y_2^2 + 2 \sum_{j \neq 2} x_2 x_j y_2 y_j$

if we found g_5 s.t.$x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

Suppose we are using the quadratic kernel:

$$k(x, y) = 1 + \sum_{i=1}^{n} x_i y_i + \sum_{i,j=1}^{n} x_i x_j y_i y_j$$

Observe $k = k_2 + k_{-2}$

- k_2 depends on x_2: $k_2(x, y) = x_2 y_2 + x_2^2 y_2^2 + 2 \sum_{j \neq 2} x_2 x_j y_2 y_j$
- k_{-2} does not depend on x_2: $k_{-2} = k - k_2$

Identify the least important ancestor

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then there is

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2

Identify the least important ancestor

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

such that

$$f_2, f_{-2} = \underset{h_2, h_{-2} \in \mathcal{H}_{k_2} \times \mathcal{H}_{k_{-2}}}{\arg\min} \begin{cases} \|h_2\|_{k_2}^2 + \|h_{-2}\|_{k_{-2}}^2 \\ \text{s.t. } h_2 + h_{-2} = g_5 \end{cases}$$

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

Then

$$g_5 = f_2 + f_{-2}$$
$$s = \|g_5\|_k^2 = \|f_2\|_{k_2}^2 + \|f_{-2}\|_{k_{-2}}^2$$

if we found g_5 s.t. $x_5 = g_5(x_1, x_2, x_3, x_4)$ using a kernel k.

$$k = k_2 + k_{-2}$$

Then there is

- $f_2 \in \mathcal{H}_{k_2}$, depends on x_2
- $f_{-2} \in \mathcal{H}_{k_{-2}}$, does not depend on x_2

Then

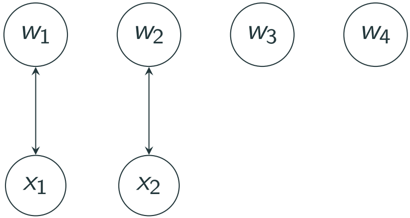
$$g_5 = f_2 + f_{-2}$$
$$s = \|g_5\|_k^2 = \|f_2\|_{k_2}^2 + \|f_{-2}\|_{k_{-2}}^2$$

We can define the activation a_2, which quantifies the contribution of x_2 to the signal data variance:

$$a_2 = \frac{\|f_2\|_{k_2}^2}{\|g_5\|_k^2}$$

Sample $w_i \sim \mathcal{N}(0,1)$ and recover the functional dependencies for

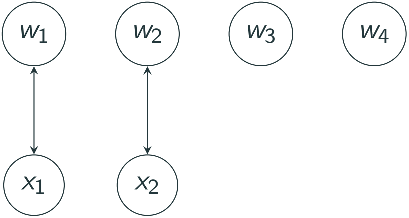
- $x_1 = w_1$ and $x_2 = w_2$



Sample $w_i \sim \mathcal{N}(0, 1)$ and recover the functional dependencies for

- $x_1 = w_1$ and $x_2 = w_2$

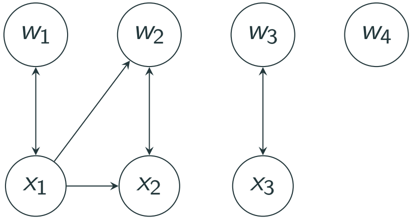
Perfect recovery with linear kernel



Algebraic equations

Sample $w_i \sim \mathcal{N}(0, 1)$ and recover the functional dependencies for

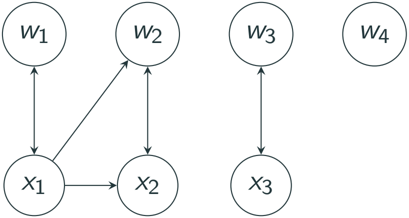
- $x_1 = w_1$ and $x_2 = w_2$
- $x_1 = w_1$,
  $x_2 = x_1^2 + 1 + 0.1 w_2$ and
  $x_3 = w_3$



Sample $w_i \sim \mathcal{N}(0,1)$ and recover the functional dependencies for

- $x_1 = w_1$ and $x_2 = w_2$
- $x_1 = w_1$,
  $x_2 = x_1^2 + 1 + 0.1w_2$ and
  $x_3 = w_3$

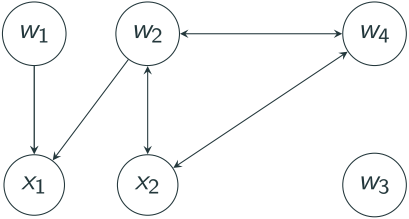
Perfect recovery with quadratic kernel



Caltech

Sample $w_i \sim \mathcal{N}(0,1)$ and recover the functional dependencies for

- $x_1 = w_1$ and $x_2 = w_2$
- $x_1 = w_1$,
  $x_2 = x_1^2 + 1 + 0.1w_2$ and
  $x_3 = w_3$
- $x_1 = w_1 w_2$, $x_2 = w_2 \sin(w_4)$

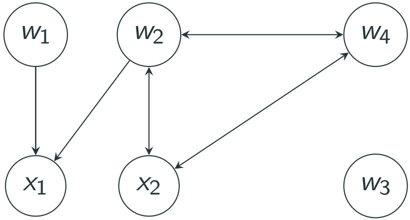


Caltech

Sample $w_i \sim \mathcal{N}(0,1)$ and recover the functional dependencies for

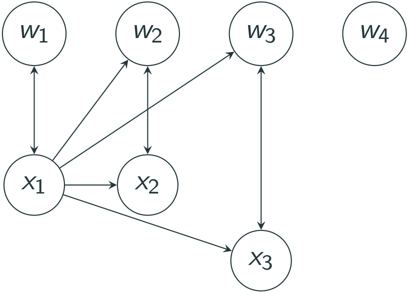
- $x_1 = w_1$ and $x_2 = w_2$
- $x_1 = w_1$, $x_2 = x_1^2 + 1 + 0.1 w_2$ and $x_3 = w_3$
- $x_1 = w_1 w_2$, $x_2 = w_2 \sin(w_4)$

Perfect recovery with non-linear kernel



Sample $w_i \sim \mathcal{N}(0, 1)$ and recover the functional dependencies for

- $x_1 = w_1$ and $x_2 = w_2$
- $x_1 = w_1$,
  $x_2 = x_1^2 + 1 + 0.1w_2$ and
  $x_3 = w_3$
- $x_1 = w_1 w_2$, $x_2 = w_2 \sin(w_4)$
- $x_1 = w_1$,
  $x_2 = x_1^3 + 1 + 0.1w_2$ and
  $x_3 = (x_1 + 2)^3 + 0.1w_3$

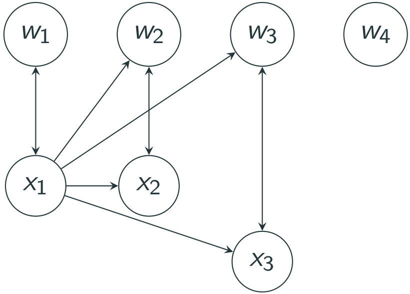


Algebraic equations

Sample $w_i \sim \mathcal{N}(0,1)$ and recover the functional dependencies for

- $x_1 = w_1$ and $x_2 = w_2$
- $x_1 = w_1$,
  $x_2 = x_1^2 + 1 + 0.1w_2$ and
  $x_3 = w_3$
- $x_1 = w_1 w_2$, $x_2 = w_2 \sin(w_4)$
- $x_1 = w_1$,
  $x_2 = x_1^3 + 1 + 0.1w_2$ and
  $x_3 = (x_1 + 2)^3 + 0.1w_3$

Unexpected perfect recovery with quadratic kernel



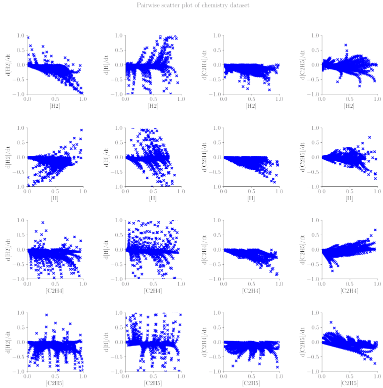
Hydrogenation of ethylene to ethane

$$C_2H_4 + H_2 \longrightarrow C_2H_6$$

$$H_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2\,H$$

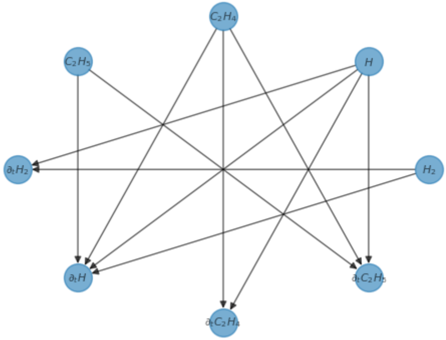
$$C_2H_4 + H \xrightarrow{k_2} C_2H_5$$

$$C_2H_5 + H \xrightarrow{k_3} C_2H_6$$



Hydrogenation of ethylene to ethane

Caltech



$$\frac{d[H_2]}{dt} = -k_1[H_2] + k_{-1}[H]^2$$
$$\frac{d[H]}{dt} = 2k_1[H_2] - 2k_{-1}[H]^2 - k_2[C_2H_4][H] - k_3[C_2H_5][H]$$
$$\frac{d[C_2H_4]}{dt} = -k_2[C_2H_4][H]$$
$$\frac{d[C_2H_5]}{dt} = k_2[C_2H_4][H] - k_3[C_2H_5][H]$$

Let $N = 10$ masses, for $i = 0, .., N-1$, their displacement from equilibrium x_i. We have:

$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2) \qquad (2)$$

Boundary condition: $x_{-1} = x_N = 0$

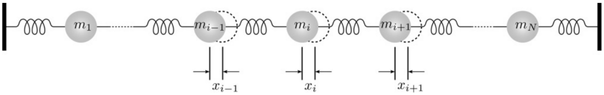


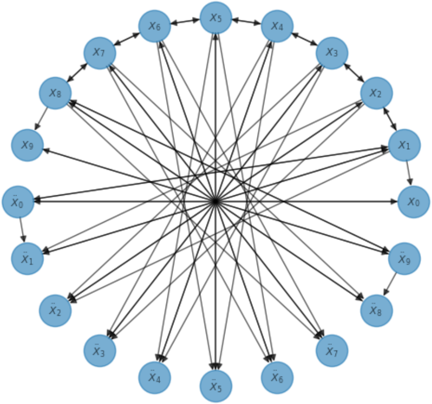
Figure 2: Nelson et al., 2018

$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2) \tag{3}$$

We observe $n = 1000$ snapshots of $x_i, \dot{x}_i, \ddot{x}_i, i = 0, .., 9$.

$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2) \tag{3}$$

We observe $n = 1000$ snapshots of $x_i, \dot{x}_i, \ddot{x}_i, i = 0, .., 9$. We recover the graph perfectly, even with uninformative prior:



A typical evolution of the noise (for $\ddot{x}_7$):

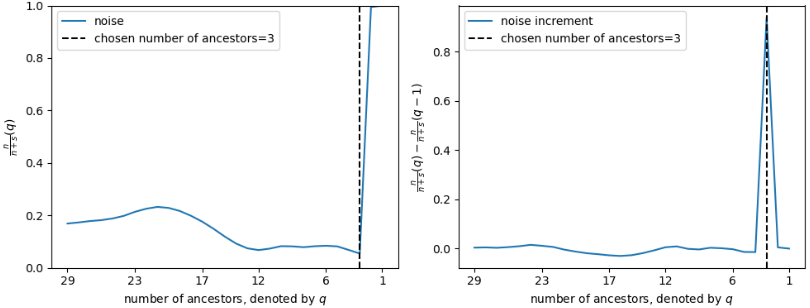


Figure 3: Left: evolution of noise-to-signal ratio . **Right**: Increment in noise ($\frac{n}{n+s}(q) - \frac{n}{n+s}(q-1)$ for q the number of ancestors)

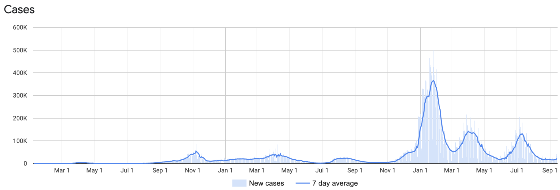
COVID dataset

Caltech

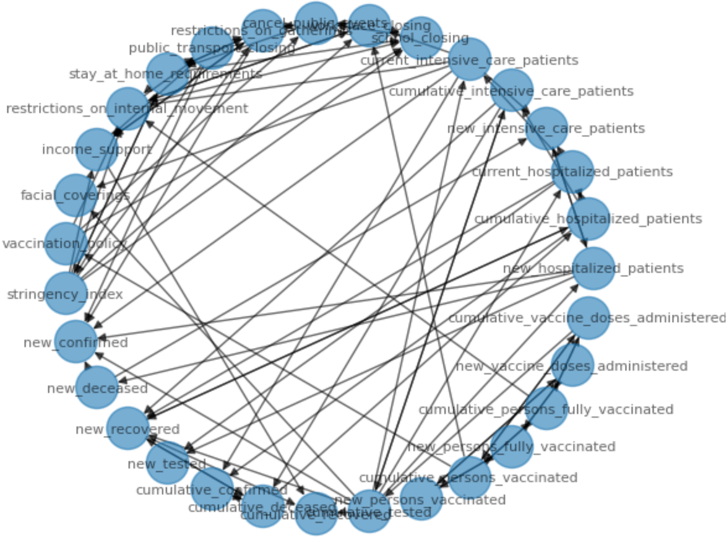
The dataset: Google's COVID data on France

Daily values of 31 variables during 500 days:

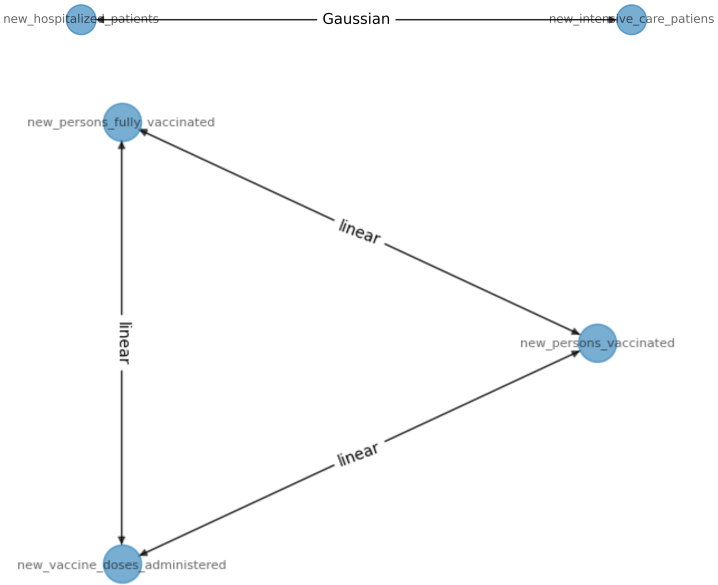
- Epidemiology dataset (new infections, cumulative deaths,...)
- Hospital dataset (number of admitted patients, patients in intensive care, etc.)
- Vaccine dataset (number of vaccinated individuals,...)
- Policy dataset (indicators related to government responses: school closures, lockdown measures, etc.)



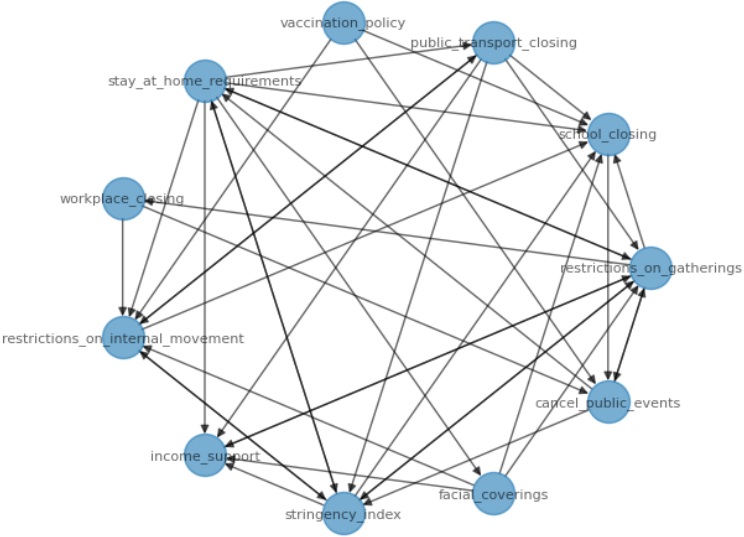
COVID dataset



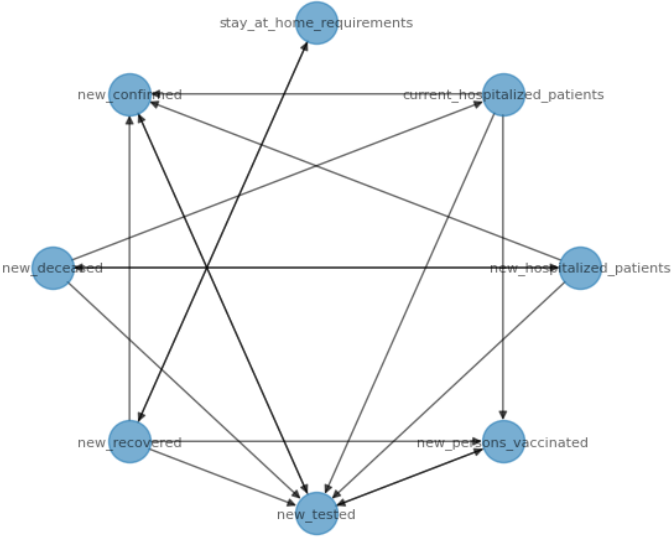
COVID dataset



COVID dataset



COVID dataset



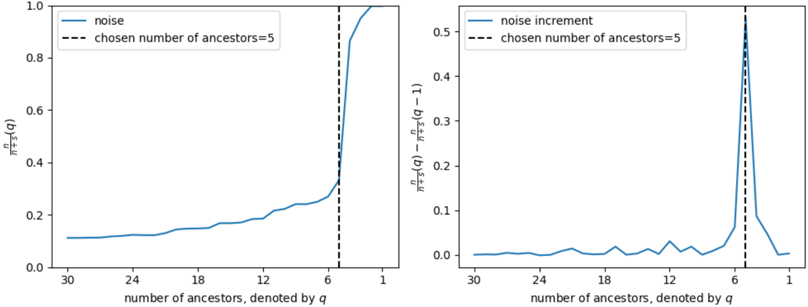


Figure 4: Evolution of the noise-to-signal ratio when pruning ancestors for the cumulative number of hospitalized patients.

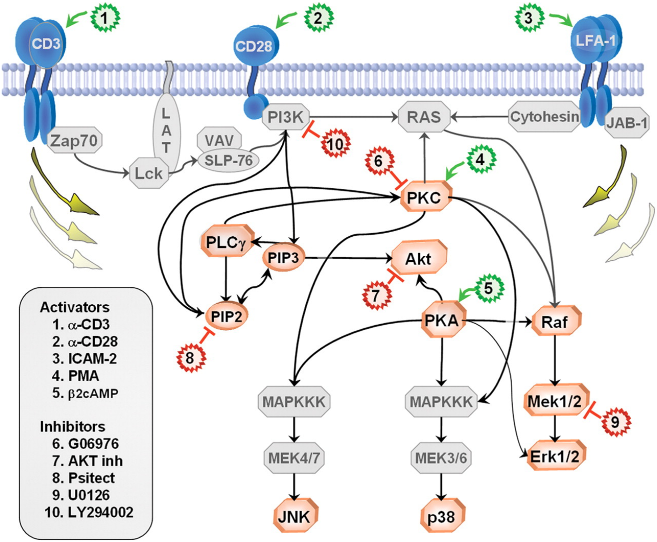
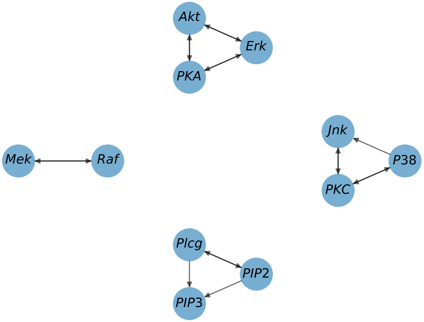


Figure 5: Sachs et al, 2005

In this dataset,

- some variables are strongly dependent, with linear dependencies (apart from 1 quadratic).
- while other dependencies are weaker and nonlinear.

To tackle this disparity, we cluster the nodes:

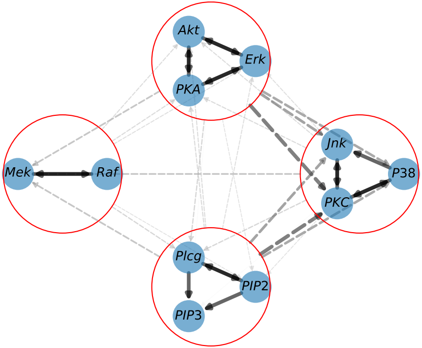


Caltech

In this dataset,

- some variables are strongly dependent, with linear dependencies (apart from 1 quadratic).
- while other dependencies are weaker and nonlinear.

To tackle this disparity, we cluster the nodes:



The Sachs dataset

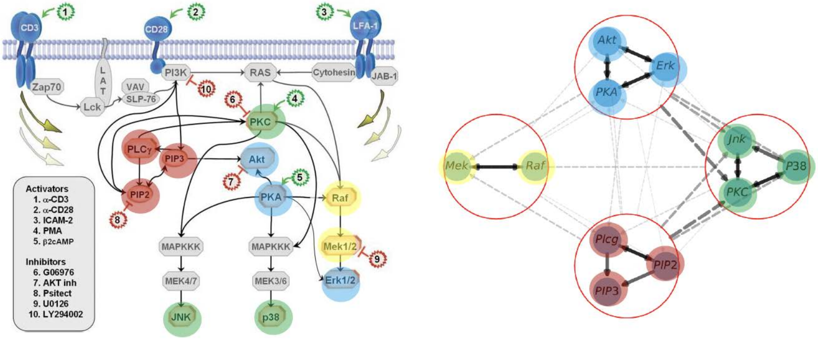


Figure 7: Comparison of recovered graph and protein signaling network

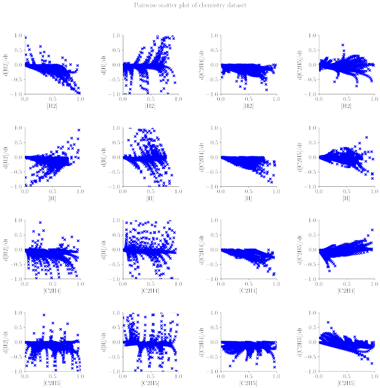
Hydrogenation of ethylene to ethane

$$C_2H_4 + H_2 \longrightarrow C_2H_6$$

$$H_2 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} 2\,H$$

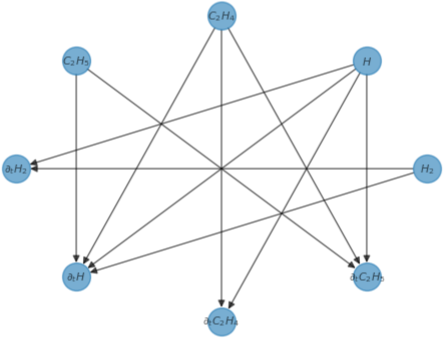
$$C_2H_4 + H \xrightarrow{k_2} C_2H_5$$

$$C_2H_5 + H \xrightarrow{k_3} C_2H_6$$



Hydrogenation of ethylene to ethane

Caltech



$$\frac{d[H_2]}{dt} = -k_1[H_2] + k_{-1}[H]^2$$

$$\frac{d[H]}{dt} = 2k_1[H_2] - 2k_{-1}[H]^2 - k_2[C_2H_4][H] - k_3[C_2H_5][H]$$

$$\frac{d[C_2H_4]}{dt} = -k_2[C_2H_4][H]$$

$$\frac{d[C_2H_5]}{dt} = k_2[C_2H_4][H] - k_3[C_2H_5][H]$$

Chemistry - Same but big

With the same process, we generate data for a larger chemical reaction. Key characteristics:

- 1122 different species.

Chemistry - Same but big

With the same process, we generate data for a larger chemical reaction. Key characteristics:

- 1122 different species.
- 5 trajectories of evolutions of the reaction are simulated.

With the same process, we generate data for a larger chemical reaction. Key characteristics:

- 1122 different species.
- 5 trajectories of evolutions of the reaction are simulated.
- We get 2400 snapshots of the system.

With the same process, we generate data for a larger chemical reaction. Key characteristics:

- 1122 different species.
- 5 trajectories of evolutions of the reaction are simulated.
- We get 2400 snapshots of the system.
- Functional dependencies are quadratic

With the same process, we generate data for a larger chemical reaction. Key characteristics:

- 1122 different species.
- 5 trajectories of evolutions of the reaction are simulated.
- We get 2400 snapshots of the system.
- Functional dependencies are quadratic so the polynomials have 600.000 coefficients for 2400 data points

Chemistry - Same but big

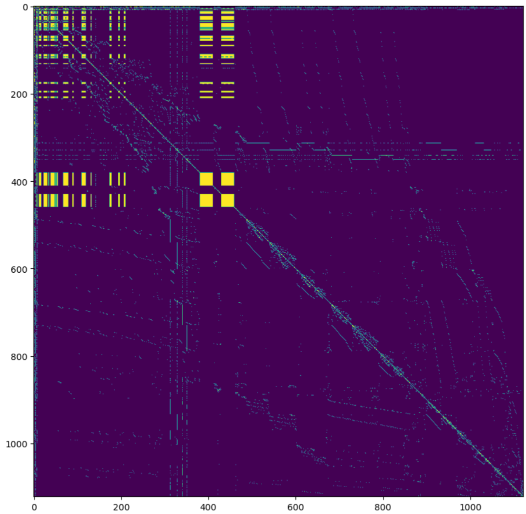


Figure 8: Adjacency matrix of the true computational hypergraph

Because we don't explore the whole phase space (only 5 initial conditions), we wouldn't be able to get this graph even if there wasn't dimensionality issues. But:

Because we don't explore the whole phase space (only 5 initial conditions), we wouldn't be able to get this graph even if there wasn't dimensionality issues. But:

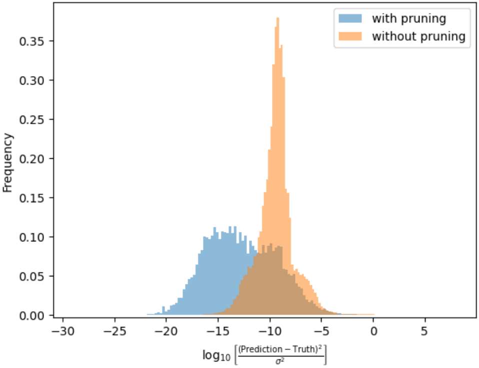


Figure 9: Improvement in accuracy after pruning

When we explore the whole phase space, dimension issues prevent us from getting the true graph. But

When we explore the whole phase space, dimension issues prevent us from getting the true graph. But

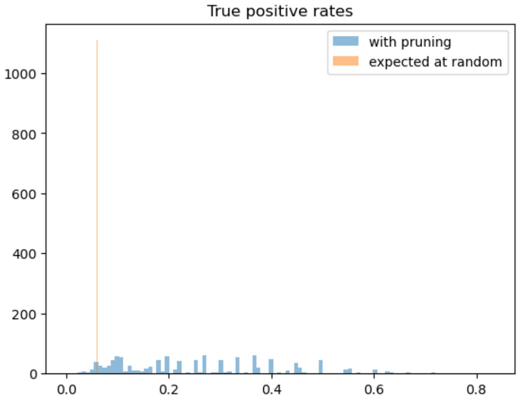


Figure 10: Histogram of percentages of true ancestors found for each nodes

Contributions

We developed a Gaussian Process-based framework to recover functional dependencies between variables

- Works for any unlabelled dataset, with few assumptions
- interpretable
- Recovers known equations in toy examples
- Yields plausible results for real datasets

Questions?

Caltech

Bourdais, T., Batlle, P., Yang, X., Baptista, R., Rouquette, N., Owhadi, H. (2024). **Codiscovering graphical structure and functional relationships within data: A Gaussian Process framework for connecting the dots.** Proceedings of the National Academy of Sciences, 121(32), e2403449121.

ComputationalHypergraphDiscovery

`pip install ComputationalHypergraphDiscovery`

Blog post on my website

Bonus

γ is a prior on the amount of noise, it influences the noise-to-signal ratio.

$$\frac{n}{n+s} = \frac{Y^T D_\gamma^2 Y}{Y^T D_\gamma Y} \text{ with } D_\gamma = \gamma(K(X,X) + \gamma I)^{-1}$$

How do you choose γ?

- If $Det(K(X,X)) = 0$ ('small' feature map dimension, $K(X,X) = \Psi(X)\Psi(X)^T$), use the residuals! (drawing inspiration from Ordinary Least Squares)

$$\gamma = \|Y - (\Psi(X)^T\Psi(X))^{-1}\Psi(X)^T Y\|^2$$

- If $Det(K(X,X)) > 0$ (kernel is interpolatory), the residuals are always 0! Calibrate γ

Proposition for $D_\gamma = \gamma(K(X,X) + \gamma I)^{-1}$

$$\lim_{\gamma \to 0} D_\gamma = 0 \preccurlyeq D_\gamma \preccurlyeq I = \lim_{\gamma \to \infty} D_\gamma$$

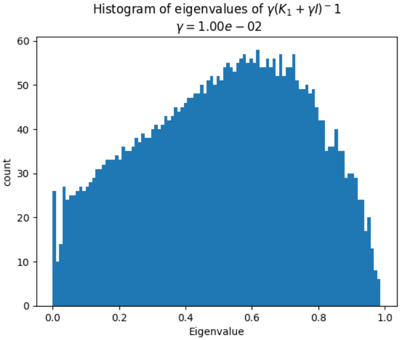


Figure 11: Good choice of γ

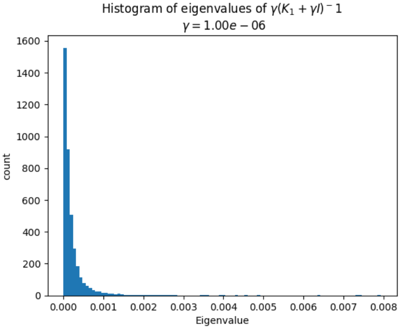


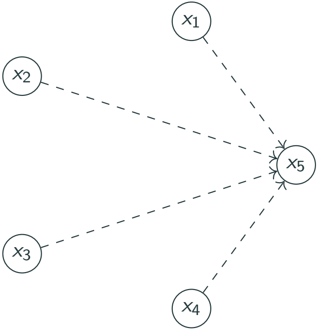
Figure 12: Bad choice of γ

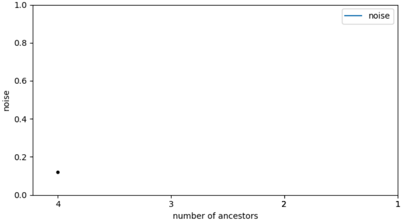
Pruning process

Caltech

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.



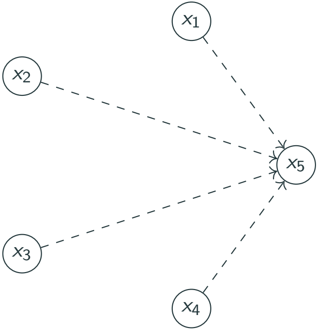


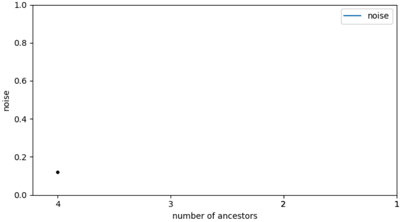
Caltech

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors



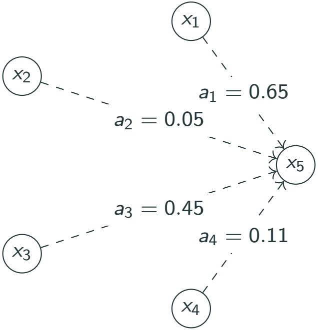


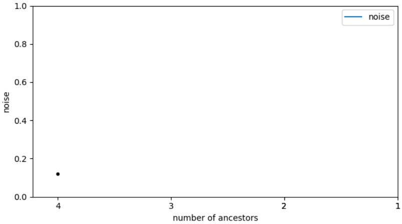
Pruning process

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

Compute initial $\frac{n}{n+s}$
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

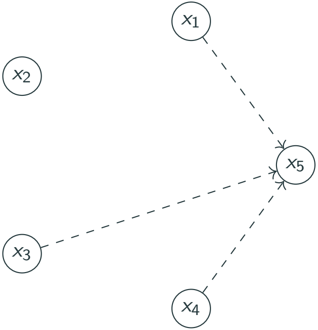


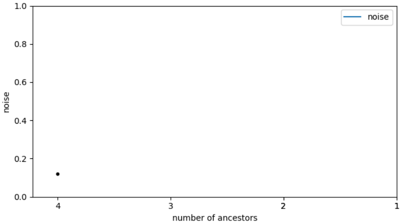


Pruning of the ancestors of x_5

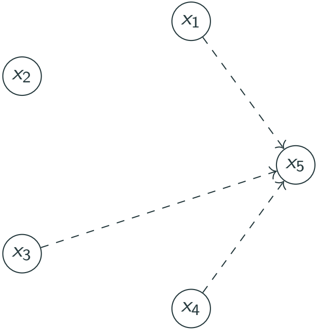
Suppose x_2 has ancestors for a kernel k.

Compute initial $\frac{n}{n+s}$
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors





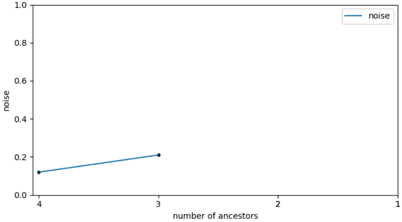
Pruning process



Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

```
Compute initial n/(n+s)
while there are some ancestors left do
    compute the contribution of each node
    remove the node that contributes the least
    recompute n/(n+s)
end while
using the evolution of n/(n+s), choose the number of ancestors
```

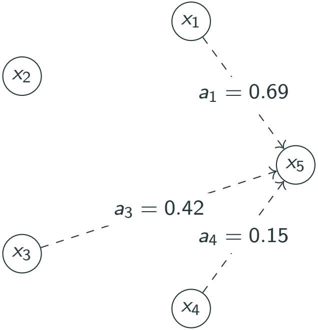


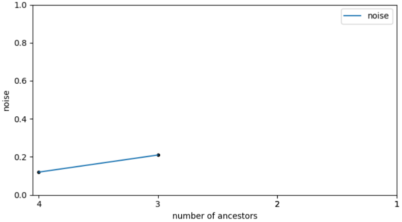
Caltech

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

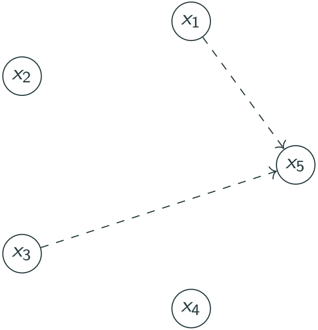


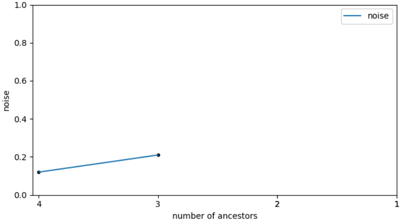


Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors



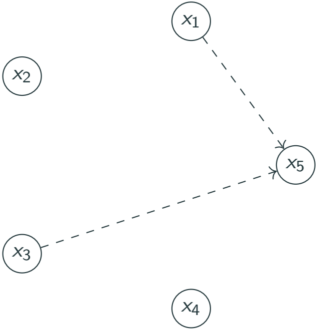


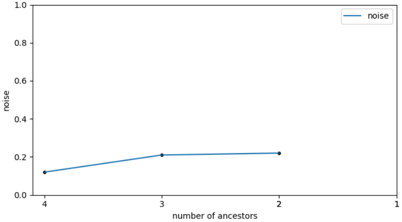
Caltech

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors





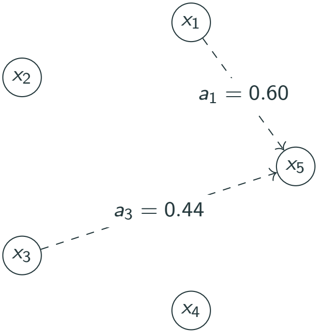
Pruning process

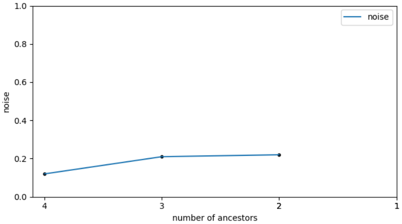
Caltech

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors



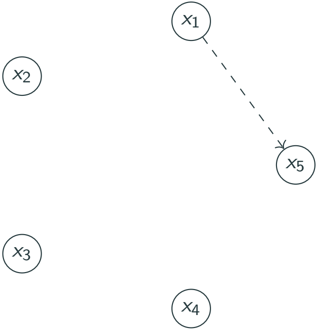


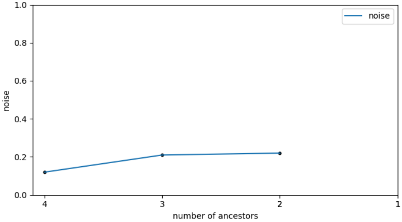
Pruning process

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors

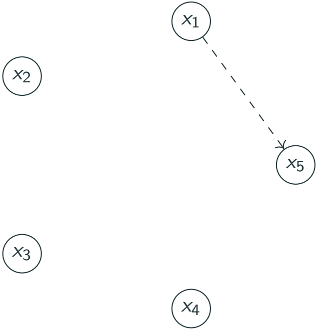


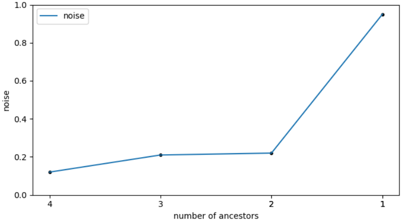


Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

Compute initial $\frac{n}{n+s}$
while there are some ancestors left **do**
    compute the contribution of each node
    remove the node that contributes the least
    recompute $\frac{n}{n+s}$
end while
using the evolution of $\frac{n}{n+s}$, choose the number of ancestors



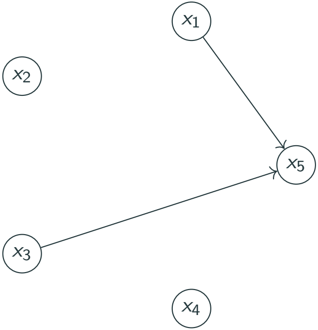


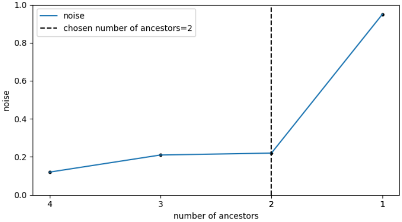
Caltech

Pruning of the ancestors of x_5

Suppose x_2 has ancestors for a kernel k.

1: Compute initial $\frac{n}{n+s}$
2: **while** there are some ancestors left **do**
3: compute the contribution of each node
4: remove the node that contributes the least
5: recompute $\frac{n}{n+s}$
6: **end while**
7: using the evolution of $\frac{n}{n+s}$, choose the number of ancestors





Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Then $\xi = \xi_2 + \xi_{-2} \sim \mathcal{N}(0, k_2 + k_{-2})$. Let

Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Then $\xi = \xi_2 + \xi_{-2} \sim \mathcal{N}(0, k_2 + k_{-2})$. Let

- $g_5 = \mathbb{E}\left[\xi \,|\xi(X) = Y\right]$: Gaussian Process estimate

Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Then $\xi = \xi_2 + \xi_{-2} \sim \mathcal{N}(0, k_2 + k_{-2})$. Let

- $g_5 = \mathbb{E}\left[\xi \,|\xi(X) = Y\right]$: Gaussian Process estimate
- $f_2 = \mathbb{E}\left[\xi_2 \,|\xi(X) = Y\right]$: contribution of k_2 (depends on x_2)

Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Then $\xi = \xi_2 + \xi_{-2} \sim \mathcal{N}(0, k_2 + k_{-2})$. Let

- $g_5 = \mathbb{E}\left[\xi \,|\xi(X) = Y\right]$: Gaussian Process estimate
- $f_2 = \mathbb{E}\left[\xi_2 \,|\xi(X) = Y\right]$: contribution of k_2 (depends on x_2)
- $f_{-2} = \mathbb{E}\left[\xi_{-2} \,|\xi(X) = Y\right]$: contribution of k_{-2} (does not depend on x_2)

Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Then $\xi = \xi_2 + \xi_{-2} \sim \mathcal{N}(0, k_2 + k_{-2})$. Let

- $g_5 = \mathbb{E}\left[\xi \,|\xi(X) = Y\right]$: Gaussian Process estimate
- $f_2 = \mathbb{E}\left[\xi_2 \,|\xi(X) = Y\right]$: contribution of k_2 (depends on x_2)
- $f_{-2} = \mathbb{E}\left[\xi_{-2} \,|\xi(X) = Y\right]$: contribution of k_{-2} (does not depend on x_2)
- $g_5 = \mathbb{E}\left[\xi_2 + \xi_{-2} \,|\xi(X) = Y\right]$

Let two independant GP:

$$\xi_2 \sim \mathcal{N}(0, k_2)$$
$$\xi_{-2} \sim \mathcal{N}(0, k_{-2})$$

Then $\xi = \xi_2 + \xi_{-2} \sim \mathcal{N}(0, k_2 + k_{-2})$. Let

- $g_5 = \mathbb{E}\left[\xi \,|\xi(X) = Y\right]$: Gaussian Process estimate
- $f_2 = \mathbb{E}\left[\xi_2 \,|\xi(X) = Y\right]$: contribution of k_2 (depends on x_2)
- $f_{-2} = \mathbb{E}\left[\xi_{-2} \,|\xi(X) = Y\right]$: contribution of k_{-2} (does not depend on x_2)
- $g_5 = \mathbb{E}\left[\xi_1 + \xi_2 \,|\xi(X) = Y\right] = f_1 + f_2$

How do you learn $x_1 = w_1$, $x_2 = x_1^3 + 1 + 0.1w_2$ and $x_3 = (x_1 + 2)^3 + 0.1w_3$ without cubic functions?

1. Use non-linear kernel to find $\tilde{f} \approx x \mapsto x^3$, or
2. Reorganise the equations into

$$\begin{aligned} x_1 &= w_1 \\ x_2 &= x_3 - 6x_1^2 - 12x_1 - 0.1w_3 + 0.1w_2 - 7 \end{aligned}$$

Here we can use only quadratic functions!