Directed dependency graph obtained from a continuous data matrix by the highest successive conditionings method

Ait-taleb Nabil*

December 23, 2021

Abstract

In this paper, we propose a directed dependency graph learned from a continuous data matrix in order to extract the hidden oriented dependencies from this matrix. For each of the dependency graph’s node, we will assign a random variable as well as a conditioning percentage linking parents and children nodes of the graph. Among all the dependency graphs learned from the continuous data matrix, we will choose the one using the highest successive conditionings method.

*Corresponding author: nabiltravail1982@gmail.com
1 Introduction

In this report, we will start by reminding the notion of differential entropy, which is useful to our learning algorithm, and used in information theory. We will apply this notion to the gaussian multidimensional probability. Later on, we will show an inequality theorem on the differential entropies in order to introduce the general and gaussian conditioning percentage. From this conditioning, we will define the concept of directed dependency graph to which we will allocate, for each node, a random variable and conditioning percentage from a current child node given the parents nodes.

Among all the dependency graphs learned from the continuous data matrix, we will choose the one using the highest successive conditionings method for each node.

The report will end with the learning of a directed dependency graph from a continuous data matrix. We will detail step by step and for each node the different iterations of the learning algorithm.
2 Information theory and Gaussian multidimensional probability

2.1 Differential entropy for a random vector

Definition: Given a random vector $\tilde{x}$, defined on set $\mathcal{X}$ of size $n$, with a multidimensional probability density function (pdf) $p_X(\tilde{x})$, we define the differential entropy $h(X)$ as:

$$h(X) = - \int_{\mathcal{X}} p_X(\tilde{x}) \ln p_X(\tilde{x}) d\tilde{x}$$

2.2 Joint differential entropy of two random vectors

Definition: Given two concatenated random vectors $(\tilde{x}_1, \tilde{x}_2)$, defined on the sets $\mathcal{X}_1$ and $\mathcal{X}_2$ of sizes $n$ and $m$ respectively, with a multidimensional probability density function (pdf) $p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2)$, we define the joint differential entropy $h(X_1, X_2)$ as:

$$h(X_1, X_2) = - \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2) \ln p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2) d\tilde{x}_1 d\tilde{x}_2$$

2.3 Conditional differential entropy of a random vector given a random vector

Definition: Given two concatenated random vectors $(\tilde{x}_1, \tilde{x}_2)$, defined on the sets $\mathcal{X}_1$ and $\mathcal{X}_2$ of sizes $n$ and $m$ respectively, with a multidimensional probability density function (pdf) $p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2)$, we define the conditional differential entropy $h(X_1|X_2)$ as:

$$h(X_1|X_2) = - \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} p_{X_1|X_2}(\tilde{x}_1, \tilde{x}_2) \ln p_{X_1|X_2}(\tilde{x}_1, \tilde{x}_2) d\tilde{x}_1 d\tilde{x}_2$$

where we have:

$$p_{X_1|X_2}(\tilde{x}_1, \tilde{x}_2) = \frac{p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2)}{p_{X_2}(\tilde{x}_2)}$$

$$p_{X_2}(\tilde{x}_2) = \int_{\mathcal{X}_1} p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2) d\tilde{x}_1$$
2.4 Joint differential entropy and conditional differential entropy

Given two concatenated random vectors \((\tilde{x}_1, \tilde{x}_2)\), defined on the sets \(X_1\) and \(X_2\) of sizes \(n\) et \(m\) respectively, with a multidimensional probability density function (pdf) \(p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2)\), we can then establish the relation between joint differential entropy and conditional differential entropy as:

\[
 h(X_1|X_2) = h(X_1, X_2) - h(X_2)
\]

Indeed:

\[
 h(X_1|X_2) = - \int_{X_1} \int_{X_2} p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2) \ln p_{X_1|X_2}(\tilde{x}_1|\tilde{x}_2) d\tilde{x}_1 d\tilde{x}_2
\]

\[
 = - \int_{X_1} \int_{X_2} p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2) \ln \frac{p_{X_1X_2}(\tilde{x}_1, \tilde{x}_2)}{p_{X_2}(\tilde{x}_2)} d\tilde{x}_1 d\tilde{x}_2
\]

\[
 = - \int_{X_1} \int_{X_2} P_{X_1|X_2}(\tilde{x}_1, \tilde{x}_2) \ln P_{X_1X_2}(\tilde{x}_1, \tilde{x}_2) d\tilde{x}_1 d\tilde{x}_2 + \int_{X_2} \left( \int_{X_1} P_{X_1|X_2}(\tilde{x}_1, \tilde{x}_2) d\tilde{x}_1 \right) \ln P_{X_2}(\tilde{x}_2) d\tilde{x}_2
\]

\[
 = h(X_1, X_2) - h(X_2)
\]
2.5 Joint and conditional gaussian multidimensional probability

Consider a partitioned random vector \( \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \) of size \( n = k_1 + k_2 \), where \( k_1 \) and \( k_2 \) are the sizes of vectors \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) respectively, with a multivariate Gaussian distribution \( P_X(\mathbf{x}) \) with a mean vector \( \mu_X \) and covariance matrix \( \Sigma_X \):

\[
P_X(\mathbf{x}) = \mathcal{N}(\mu_X, \Sigma_X) = (2\pi)^{-\frac{\mathbf{x}}{2}}|\Sigma_X|^{-\frac{1}{2}} \exp\left(-\frac{(\mathbf{x} - \mu_X)'\Sigma_X^{-1}(\mathbf{x} - \mu_X)}{2}\right)
\]

The purpose of this section is to expose the following different probabilities:

1. \( P_X(\mathbf{x}) = P_{X_1,X_2}(\mathbf{x}_1, \mathbf{x}_2) \)
2. \( P_{X_1}(\mathbf{x}_2) \)
3. \( P_{X_1|X_2}(\mathbf{x}_1, \mathbf{x}_2) \)

For this, we must start first from the block matrix multiplication of the covariance matrix \( \Sigma \) and the precision matrix \( \Sigma^{-1} \) and prove the following relation:

\[
W_{X_1} = \Sigma^{-1} X_1 + W_{X_1} \Sigma^{-1} \times X_1.
\]

Indeed:

\[
K_{X_1} W_{X_1} = \begin{pmatrix} K_{X_1} W_{X_1} + K_{X_1} W_{X_1} & K_{X_1} W_{X_1} + K_{X_1} W_{X_1} \end{pmatrix}
\]

\[
K_{X_1} W_{X_1} + K_{X_1} W_{X_1} = 0
\]

\[
K_{X_1} W_{X_1} + K_{X_1} W_{X_1} = 0
\]

\[
K_{X_1} W_{X_1} = -W_{X_1} W_{X_1}^{-1}
\]

\[
K_{X_1} W_{X_1} + K_{X_1} W_{X_1} = I_{k_1}
\]

\[
W_{X_1} = K_{X_1}^{-1} \times K_{X_1} + W_{X_1} W_{X_1}^{-1} \times W_{X_1}
\]

Finally, we obtain:

\[
W_{X_1} = K_{X_1}^{-1} + W_{X_1} W_{X_1}^{-1} \times W_{X_1}
\]

Now, we will develop the Mahalanobis distance:

\[
(\mathbf{x} - \mu_X)' W_X (\mathbf{x} - \mu_X)
\]

\[
= (\mathbf{x}_1 - \mu_X_1, \mathbf{x}_2 - \mu_X_2)' \begin{pmatrix} W_{X_1} & W_{X_1} W_{X_1}^{-1} \mathbf{x}_1 - \mu_X_1 \\ W_{X_1} W_{X_1}^{-1} \mathbf{x}_2 - \mu_X_2 \end{pmatrix}
\]

\[
= (\mathbf{x}_1 - \mu_X_1)' W_{X_1} (\mathbf{x}_1 - \mu_X_1) + (\mathbf{x}_1 - \mu_X_1)' W_{X_1} (\mathbf{x}_1 - \mu_X_1) + (\mathbf{x}_2 - \mu_X_2)' W_{X_1} (\mathbf{x}_2 - \mu_X_2)
\]

\[
+ (\mathbf{x}_2 - \mu_X_2)' W_{X_1} (\mathbf{x}_2 - \mu_X_2)
\]

Using the relation: \( W_{X_1} = K_{X_1}^{-1} + W_{X_1} W_{X_1}^{-1} \times W_{X_1} \), we obtain:
\[(\bar{x}_1 - \mu_{\bar{x}_1})' W_{X_1'} (\bar{x}_1 - \mu_{\bar{x}_1}) + (\bar{x}_1 - \mu_{\bar{x}_1})' W_{X_1'} (\bar{x}_2 - \mu_{\bar{x}_2}) + (\bar{x}_2 - \mu_{\bar{x}_2})' W_{X_1'} (\bar{x}_1 - \mu_{\bar{x}_1}) + (\bar{x}_2 - \mu_{\bar{x}_2})' W_{X_1'} (\bar{x}_2 - \mu_{\bar{x}_2})
\]
\[= [(\bar{x}_1 - \mu_{\bar{x}_1}) + W_{X_1'}^{-1} W_{X_1' X_2} (\bar{x}_2 - \mu_{\bar{x}_2})]' [W_{X_1'} (\bar{x}_1 - \mu_{\bar{x}_1}) + W_{X_1' X_2} (\bar{x}_2 - \mu_{\bar{x}_2})]
\]
\[+ (\bar{x}_2 - \mu_{\bar{x}_2})' K_{X_2}^{-1} (\bar{x}_2 - \mu_{\bar{x}_2})
\]
\[= [(\bar{x}_1 - \mu_{\bar{x}_1}) + W_{X_1'}^{-1} W_{X_1' X_2} (\bar{x}_2 - \mu_{\bar{x}_2})]' W_{X_1'} [(\bar{x}_1 - \mu_{\bar{x}_1}) + W_{X_1' X_2} (\bar{x}_2 - \mu_{\bar{x}_2})]
\]
\[+ (\bar{x}_2 - \mu_{\bar{x}_2})' K_{X_2}^{-1} (\bar{x}_2 - \mu_{\bar{x}_2})
\]

We then obtain the equalities as follows:

We put:

\[Q_1 = (\bar{x}_1 - \nu_{X_1'} X_1) (K_{X_1' X_1} - K_{X_1' X_2} K_{X_2}^{-1} K_{X_2' X_1})^{-1} (\bar{x}_1 - \nu_{X_1'} X_1)
\]
\[\nu_{X_1'} X_1 = \mu_{\bar{x}_1} + K_{X_1' X_1} K_{X_2}^{-1} (\bar{x}_2 - \mu_{\bar{x}_2})
\]
\[Q_2 = (\bar{x}_2 - \mu_{\bar{x}_2})' K_{X_2}^{-1} (\bar{x}_2 - \mu_{\bar{x}_2})
\]

We then obtain the equalities as follows:

\[(\bar{x} - \mu_{\bar{x}})' K_{\bar{x}}^{-1} (\bar{x} - \mu_{\bar{x}}) = Q_1 + Q_2
\]
\[P_{X_1} (\bar{x}) = (2\pi)^{-\frac{1}{2}} |K_{X_1'}|^{-\frac{1}{2}} \exp \left( - \frac{\nu_{\bar{x}_1'} X_1' - \frac{1}{2} \nu_{\bar{x}_1'} X_1' \nu_{\bar{x}_1'} X_1' \nu_{\bar{x}_1'} X_1' |K_{X_1'}|^{-\frac{1}{2}}(\bar{x} - \mu_{\bar{x}}) \right)
\]
\[P_{X_2} (\bar{x}_2) = (2\pi)^{-\frac{1}{2}} |K_{X_2'}|^{-\frac{1}{2}} \exp \left( - \frac{\nu_{\bar{x}_2'} X_2' - \frac{1}{2} \nu_{\bar{x}_2'} X_2' \nu_{\bar{x}_2'} X_2' |K_{X_2'}|^{-\frac{1}{2}}(\bar{x}_2 - \mu_{\bar{x}_2}) \right)
\]

Using the relation \[\frac{P_{X_1} (\bar{x})}{P_{X_2} (\bar{x}_2)}: \]
\[P_{X_1|X_2} (x_1, x_2) = (2\pi)^{-\frac{1}{2}} \frac{|K_{X_1 X_1'}|}{|K_{X_2'}|^{-\frac{1}{2}}} \exp \left( - \frac{\nu_{\bar{x}_1'} X_1' - \frac{1}{2} \nu_{\bar{x}_1'} X_1' \nu_{\bar{x}_1'} X_1' |K_{X_1'}|^{-\frac{1}{2}} (x_1 - \mu_{x_1}) - \frac{\nu_{\bar{x}_2'} X_2' - \frac{1}{2} \nu_{\bar{x}_2'} X_2' \nu_{\bar{x}_2'} X_2' |K_{X_2'}|^{-\frac{1}{2}} (x_2 - \mu_{x_2}) \right)
\]

If we use the Schur’s complement \[K_{X_1|X_2} = K_{X_1'} - K_{X_1' X_2} K_{X_2}^{-1} K_{X_2' X_1}\]
we can express the conditional probability \[P_{X_1|X_2} (\bar{x}_1, \bar{x}_2)\] as follows:

\[P_{X_1|X_2} (\bar{x}_1, \bar{x}_2) = (2\pi)^{-\frac{1}{2}} \frac{|K_{X_1 X_1'}|}{|K_{X_2'}|^{-\frac{1}{2}}} \exp \left( - \frac{(\bar{x}_1 - \nu_{\bar{x}_1'} X_1') K_{X_1'}^{-1} (\bar{x}_1 - \nu_{\bar{x}_1'} X_1')}{2} - \frac{(\bar{x}_2 - \nu_{\bar{x}_2'} X_2') K_{X_2'}^{-1} (\bar{x}_2 - \nu_{\bar{x}_2'} X_2')}{2} \right)
\]
2.6 Differential entropy of a Gaussian random vector

**Theorem:** Given random vector \( \tilde{x} = (x_1, x_2, \ldots, x_n) \) with a multivariate Gaussian distribution:

\[
P_X(\tilde{x}) = \mathcal{N}(\mu_X, K_{X^2}) = (2\pi)^{-\frac{n}{2}}|K_{X^2}|^{-\frac{1}{2}} \exp\left(-\frac{(\tilde{x} - \mu_X)^tK_{X^2}^{-1}(\tilde{x} - \mu_X)}{2}\right)
\]

with a mean vector \( \mu_X \) and a covariance matrix \( K_{X^2} \) then the differential entropy is equal to:

\[
h(X) = \frac{1}{2} \ln(2\pi e)^n|K_{X^2}|
\]

**Proof:**

\[
h(X) = -\int_{-\infty}^{+\infty} p_X(\tilde{x}) \ln(p_X(\tilde{x})) d\tilde{x}
\]

\[
= -\int_{-\infty}^{+\infty} p_X(\tilde{x}) \left[-\frac{1}{2}(\tilde{x} - \mu_X)^tK_{X^2}^{-1}(\tilde{x} - \mu_X) - \ln(\sqrt{2\pi})\right] d\tilde{x}
\]

\[
= \frac{1}{2} E_X[\sum_{ij}(\tilde{x}_i - \mu_X_i)^t(K_{X^2}^{-1})_{ij}(\tilde{x}_j - \mu_X_j)] + \frac{1}{2} \ln(2\pi)^n|K_{X^2}|
\]

\[
= \frac{1}{2} E_X[\sum_{ij}(\tilde{x}_i - \mu_X_i)^t(\tilde{x}_j - \mu_X_j)(K_{X^2}^{-1})_{ij}] + \frac{1}{2} \ln(2\pi)^n|K_{X^2}|
\]

\[
= \frac{1}{2} \sum_{ij} E_X[(K_{X^2})_{ji}(K_{X^2}^{-1})_{ij}] + \frac{1}{2} \ln(2\pi)^n|K_{X^2}|
\]

\[
= \frac{1}{2} \sum_j((K_{X^2})_{jj}(K_{X^2}^{-1})_{jj}] + \frac{1}{2} \ln(2\pi)^n|K_{X^2}|
\]

\[
= \frac{1}{2} \sum_j I_{jj} + \frac{1}{2} \ln(2\pi)^n|K_{X^2}|
\]

\[
= \frac{n}{2} + \frac{1}{2} \ln(2\pi)^n|K_{X^2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^n|K_{X^2}|
\]
2.7 Conditional differential entropy of two Gaussian random vectors

**Theorem:** Given two concatenated Gaussian random vectors \( \bar{x} = (\tilde{x}_1, \tilde{x}_2) \), of sizes \( k_1 \) and \( k_2 \) respectively, with a multivariate Gaussian distribution:

\[
P_X(\bar{x}) = \mathcal{N}(\mu_X, K_X) = (2\pi)^{-\frac{K}{2}}|K_X|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\bar{x} - \mu_X)'K_X^{-1}(\bar{x} - \mu_X)\right\}
\]

with a mean vector \( \mu_X \) and a covariance matrix \( K_X \).

In this case, the conditional differential entropy \( h(X_1|X_2) \) is equal to:

\[
h(X_1|X_2) = \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

**Proof:**

\[
h(X_1|X_2) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X_1,X_2}(\tilde{x}_1, \tilde{x}_2) \ln(p_{X_1}(\tilde{x}_1)) \frac{dx_1}{2} d\tilde{x}_2
\]

We know the conditional probability \( P_{X_1|X_2} \) can be expressed as follows:

\[
P_{X_1|X_2}(\tilde{x}_1, \tilde{x}_2) = (2\pi)^{-\frac{k_1}{2}}|K_{X_1|X_2}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\tilde{x}_1 - \mu_{X_1|X_2})'K_{X_1|X_2}^{-1}(\tilde{x}_1 - \mu_{X_1|X_2})\right\}
\]

So we can write:

\[
h(X_1|X_2) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X_1,X_2}(\tilde{x}_1, \tilde{x}_2)[-\frac{1}{2}(\tilde{x}_1 - \mu_{X_1|X_2})'K_{X_1|X_2}^{-1}(\tilde{x}_1 - \mu_{X_1|X_2}) - \ln(\sqrt{2\pi})^{k_1} |K_{X_1|X_2}|^{\frac{1}{2}}] dx_1 d\tilde{x}_2
\]

\[
= \frac{1}{2} E_{X_1,X_2}[\sum_{ij} ((\tilde{x}_i) - \mu_{X_1|X_2})'K_{X_1|X_2}^{-1}((\tilde{x}_i) - \mu_{X_1|X_2})] + \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \sum_{ij} E_{X_1,X_2}[(\tilde{x}_i) - \mu_{X_1|X_2})'K_{X_1|X_2}^{-1}((\tilde{x}_i) - \mu_{X_1|X_2})] + \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \sum_{ij} (K_{X_1|X_2})_{ij} + \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \sum_{i} I_{ii} + \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]

\[
= \frac{1}{2} \ln(2\pi e)^{k_1} |K_{X_1|X_2}|
\]
2.8 Inequalities theorem on the conditional differential entropies gaussian vectors

**Theorem:** Given a partitioned Gaussian random vector \( \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \), of sizes \( k_1 \), \( k_2 \) and \( k_3 = 1 \) respectively, with the multivariate Gaussian distribution \( \mathcal{N}(\mu_{\mathbf{x}}, K_{\mathbf{x}}) \) then we can write the following inequalities:

\[
h(X_3|X_1, X_2) \leq h(X_3|X_1) \leq h(X_3)
\]

**Proof:**

For this, we must start first from the block matrix multiplication of the covariance matrix \( K_{(X_1, X_2)} \) and the precision matrix \( W_{(X_1, X_2)} = K_{(X_1, X_2)}^{-1} \) and prove the following relation:

\[
\begin{align*}
K_{(X_1, X_2)} W_{(X_1)} & = K_{X_1}^{-1} + W_{X_1, X_2}^{-1} W_{X_2, X_1} \\
K_{(X_1, X_2)} W_{X_1} & = \begin{pmatrix} K_{X_1} & W_{X_1, X_2} \\ W_{X_2, X_1} & K_{X_2} \end{pmatrix} = \begin{pmatrix} I_{k_1, k_1} & 0 \\ 0 & I_{k_2, k_2} \end{pmatrix} \\
K_{(X_1, X_2)} & = W_{X_1}^{-1} + W_{X_2, X_1}^{-1} W_{X_2, X_1} \\
W_{X_1} & = K_{X_1}^{-1} + K_{X_1, X_2} W_{X_2, X_1} \\
W_{X_2} & = K_{X_2}^{-1} + W_{X_1, X_2}^{-1} W_{X_2, X_1}
\end{align*}
\]

We must develop the following quadratic form for \( n = k_1 + k_2 + k_3 = k_1 + k_2 + 1 \):

\[
(K_{(X_1, X_2)}, K_{(X_1, X_3)}) K_{(X_1, X_2)}^{-1} (K_{X_1, X_3}) = (K_{X_1, X_3}) W_{(X_1, X_3)}^{-1} (K_{X_1, X_3})
\]

or

\[
(K_{X_1, X_3}) W_{X_1, X_3}^{-1} W_{X_1, X_3} (K_{X_1, X_3}) = (K_{X_1, X_3}) W_{X_1, X_3}^{-1} W_{X_1, X_3} (K_{X_1, X_3})
\]

Using the relation: \( W_{X_1} = K_{X_1}^{-1} + W_{X_1, X_2} W_{X_2, X_1}^{-1} W_{X_2, X_1} \):

\[
(K_{X_1, X_3}) W_{X_1, X_3}^{-1} W_{X_1, X_3} (K_{X_1, X_3}) = (K_{X_1, X_3}) W_{X_1, X_3}^{-1} W_{X_1, X_3} (K_{X_1, X_3})
\]

9
Finally, we have the relation:

\[ h(X_3 | X_1, X_2) ≤ h(X_3 | X_1) ≤ h(X_3) \]

However both following quadratics forms are equivalents:

\[ (K_{X_1, X_2}, K_{X_1 X_2}) \begin{pmatrix} W_{X_1}^2 & W_{X_1 X_2} & K_{X_1 X_2} \\ W_{X_1 X_2} & W_{X_2}^2 & K_{X_1 X_2} \end{pmatrix} = (K_{X_1, X_2}, K_{X_1 X_2}) \begin{pmatrix} W_{X_2}^2 & W_{X_2 X_1} & K_{X_1 X_2} \\ W_{X_2 X_1} & W_{X_1}^2 & K_{X_1 X_2} \end{pmatrix} \]

and are positive semidefinite if and only if:

\[ W_{X_1}^2 ≥ 0, W_{X_2}^2 - W_{X_1 X_2}, W_{X_2 X_1} ≥ 0 \]

but yet

\[ W_{X_1}^2 ≥ 0, W_{X_2}^2 - W_{X_1 X_2}, W_{X_2 X_1} ≥ 0 \]

As \( W_{X_1}^2 ≥ 0 \), we can write the inequalities:

\[ (K_{X_1, X_2}, K_{X_1 X_2}) K_{X_1 X_2}^{-1} (K_{X_1, X_2}) ≥ 0 \]

If we use \( K_{X_2}^2 \), we can write:

\[ K_{X_2}^2 ≤ (K_{X_1, X_2}, K_{X_1 X_2}) K_{X_1 X_2}^{-1} (K_{X_1, X_2}) ≤ K_{X_2}^2 \]

\[ \frac{1}{2} \ln |K_{X_1, X_2}| + \frac{1}{2} \ln (2πe)^n ≤ \frac{1}{2} \ln |K_{X_2}| + \frac{1}{2} \ln (2πe)^n \]

Finally, we have the relation:

\[ h(X_3 | X_1, X_2) ≤ h(X_3 | X_1) ≤ h(X_3) \]
3 General conditioning percentage

We use the inequalities \( h(X_3|X_1, X_2) \leq h(X_3|X_1) \leq h(X_3) \) to define the conditioning percentage.

**Definition:** Given a set of variables \( \Omega \equiv X_{n+1,...,n} \), the variable \( X_j \in \Omega \), the subsets \( \Omega_1 \subset \Omega \setminus X_j \) and \( \Omega_2 \equiv \Omega \setminus \{X_j, \Omega_1\} \) we can define the conditioning percentage \( \lambda_{X_j|\Omega_1} \) of the variables \( \Omega_1 \) which act on the variable \( X_j \) as follows:

\[
\lambda_{X_j|\Omega_1} = \frac{h(X_j) - h(X_j|\Omega_1)}{h(X_j) - h(X_j|\Omega_1, \Omega_2)} = \frac{E_{\Omega}[\ln P_{X_j|\Omega_1}(x_j, \omega_1)]}{h(X_j) - h(X_j|\Omega_1)} + \frac{h(X_j)}{h(X_j) - h(X_j|\Omega_1)}
\]

\[0 \leq \lambda_{X_j|\Omega_1} \leq 1\]

From inequalities \( h(X_j|\Omega_1, \Omega_2) \leq h(X_j|\Omega_1) \leq h(X_j) \), if we have the equality: \( h(X_j) = h(X_j|\Omega_1, \Omega_2) \) then \( h(X_j|\Omega_1) = h(X_j) \). In this case, \( X_j \perp \Omega_1 \cup \Omega_2 \) and \( \lambda_{X_j|\Omega_1} = 0 \). (where the symbol \( \perp \) corresponds to the independency symbol)

If \( h(X_j) \neq h(X_j|\Omega_1, \Omega_2) \) and \( h(X_j) = h(X_j|\Omega_1) \) then \( X_j \perp \Omega_1 \) and \( \lambda_{X_j|\Omega_1} = 0 \)

4 Gaussian conditioning percentage

**Definition:** Given a set of Gaussian variables \( \Omega \equiv X_{n+1,...,n} \), the Gaussian variable \( X_j \in \Omega \), the subsets \( \Omega_1 \subset \Omega \setminus X_j \) and \( \Omega_2 \equiv \Omega \setminus \{X_j, \Omega_1\} \) we can define the Gaussian conditioning percentage \( \lambda_{X_j|\Omega_1} \) of the Gaussian variables \( \Omega_1 \) which act on the Gaussian variable \( X_j \) as follows:

\[
\lambda_{X_j|\Omega_1} = \frac{h(X_j) - h(X_j|\Omega_1)}{h(X_j) - h(X_j|\Omega_1, \Omega_2)} = \frac{\frac{1}{2} \ln(2\pi.e.K_{X_j}) - \frac{1}{2} \ln(2\pi.e.K_{X_j|\Omega_1})}{\frac{1}{2} \ln(2\pi.e.K_{X_j}) - \frac{1}{2} \ln(2\pi.e.K_{X_j|\Omega_1, \Omega_2})}
\]

In what follows, we will consider the gaussian entropy and the gaussian conditioning percentage for the continuous data matrix learning.
5 Directed dependency graph

Definition:
The directed dependency graph is directed graph to which we attribute for each node a random variable and the conditioning percentage linked to the edges going from the set of nodes $\Omega_1$ to the node $X_j$:

$$\lambda_{X_j|\Omega_1} = \frac{h(X_j) - h(X_j|\Omega_1)}{h(X_j) - h(X_j|\Omega_1, \Omega_2)} = \frac{E_{\Omega}[\ln P_{X_j|\Omega_1}(x_j, \omega_1)]}{h(X_j) - h(X_j|\Omega_1, \Omega_2)} + \frac{h(X_j)}{h(X_j) - h(X_j|\Omega_1, \Omega_2)}$$

If $\lambda_{X_j|\Omega_1} = 0$, $h(X_j) = h(X_j|\Omega_1)$ and $h(X_j) \neq h(X_j|\Omega_1, \Omega_2)$ then $X_j \perp \Omega_1$

If $\lambda_{X_j|\Omega_1} = 0$ and $h(X_j) = h(X_j|\Omega_1) = h(X_j|\Omega_1, \Omega_2)$ then $X_j \perp \Omega_1 \cup \Omega_2$.

6 The highest successive conditionings method

Among all the dependency graphs learned from the continuous data matrix, we will choose the one using the highest successive conditionings method for each node. The highest successive conditionings method is a method which makes it possible to obtain the smallest subsets of parent nodes which most strongly condition the child nodes. For each node $X_i$, this method can be described by the following optimization problem:

$$\max_{X_i} \lambda_{X_i|x_j} = \min_{X_j} h(X_i|x_j)$$

$$\max_{X_i} \lambda_{X_i|x_j, X_k} = \min_{X_k} h(X_i|x_j, X_k)$$

$$\max_{X_i} \lambda_{X_i|x_j, X_k, X_l} = \min_{X_l} h(X_i|x_j, X_k, X_l)$$

⋮

For this optimization problem, we will set a value to be exceeded by the conditioning percentage for each node. This value will be set to 95% ($\lambda = 0.95$), in which case, the dependencies model will be considered as good.

In what follows, we will propose the learning method applied to a continuous data matrix that contains car brands.
7 Directed dependency graph learned from continuous data matrix

7.1 Continuous data matrix

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21.697356</td>
<td>212.496303</td>
<td>100.27983</td>
<td>4.067217</td>
<td>3.112837</td>
<td>2.045330</td>
</tr>
<tr>
<td>2</td>
<td>17.933487</td>
<td>334.547171</td>
<td>216.25136</td>
<td>3.032607</td>
<td>3.945227</td>
<td>17.51779</td>
</tr>
<tr>
<td>3</td>
<td>22.593178</td>
<td>293.789279</td>
<td>131.11323</td>
<td>3.847017</td>
<td>2.874565</td>
<td>17.27431</td>
</tr>
<tr>
<td>4</td>
<td>34.049362</td>
<td>-140.459877</td>
<td>59.76671</td>
<td>5.185856</td>
<td>0.378113</td>
<td>18.91340</td>
</tr>
<tr>
<td>5</td>
<td>18.893331</td>
<td>193.854070</td>
<td>3.945227</td>
<td>1.274225</td>
<td>2.045330</td>
<td>17.51779</td>
</tr>
<tr>
<td>6</td>
<td>27.386443</td>
<td>183.449699</td>
<td>89.49365</td>
<td>4.120053</td>
<td>0.378113</td>
<td>18.91340</td>
</tr>
<tr>
<td>7</td>
<td>29.387658</td>
<td>-27.047273</td>
<td>48.51673</td>
<td>4.005462</td>
<td>1.274225</td>
<td>17.51779</td>
</tr>
<tr>
<td>8</td>
<td>13.803899</td>
<td>289.576913</td>
<td>203.10891</td>
<td>2.691737</td>
<td>4.283253</td>
<td>17.28728</td>
</tr>
<tr>
<td>9</td>
<td>23.307997</td>
<td>190.350364</td>
<td>83.14425</td>
<td>4.433126</td>
<td>2.325246</td>
<td>15.05471</td>
</tr>
<tr>
<td>10</td>
<td>34.096057</td>
<td>2.741246</td>
<td>47.94401</td>
<td>3.978904</td>
<td>0.845964</td>
<td>18.26750</td>
</tr>
<tr>
<td>12</td>
<td>12.740558</td>
<td>351.668352</td>
<td>214.58385</td>
<td>2.513282</td>
<td>4.260083</td>
<td>19.36172</td>
</tr>
<tr>
<td>13</td>
<td>19.228647</td>
<td>256.121885</td>
<td>158.85563</td>
<td>4.233123</td>
<td>2.325246</td>
<td>18.05137</td>
</tr>
<tr>
<td>14</td>
<td>25.065331</td>
<td>192.011334</td>
<td>142.97339</td>
<td>3.878321</td>
<td>2.142523</td>
<td>17.52720</td>
</tr>
<tr>
<td>15</td>
<td>26.419988</td>
<td>72.476092</td>
<td>79.17064</td>
<td>3.985763</td>
<td>2.283253</td>
<td>18.05137</td>
</tr>
<tr>
<td>16</td>
<td>9.768450</td>
<td>360.315682</td>
<td>190.86103</td>
<td>2.513282</td>
<td>4.260083</td>
<td>19.36172</td>
</tr>
<tr>
<td>17</td>
<td>17.009185</td>
<td>351.668352</td>
<td>214.58385</td>
<td>2.513282</td>
<td>4.260083</td>
<td>19.36172</td>
</tr>
<tr>
<td>19</td>
<td>23.293301</td>
<td>187.338921</td>
<td>137.30990</td>
<td>3.753809</td>
<td>2.936190</td>
<td>18.05137</td>
</tr>
<tr>
<td>20</td>
<td>24.776262</td>
<td>138.13614</td>
<td>118.17383</td>
<td>4.371926</td>
<td>1.426564</td>
<td>18.26750</td>
</tr>
<tr>
<td>21</td>
<td>22.308028</td>
<td>5.969799</td>
<td>118.17383</td>
<td>4.371926</td>
<td>1.426564</td>
<td>18.26750</td>
</tr>
<tr>
<td>22</td>
<td>19.228647</td>
<td>229.179303</td>
<td>148.90931</td>
<td>3.141172</td>
<td>3.222483</td>
<td>18.05137</td>
</tr>
<tr>
<td>23</td>
<td>25.065331</td>
<td>192.011334</td>
<td>142.97339</td>
<td>3.878321</td>
<td>2.142523</td>
<td>17.52720</td>
</tr>
<tr>
<td>24</td>
<td>26.419988</td>
<td>72.476092</td>
<td>79.17064</td>
<td>3.985763</td>
<td>2.283253</td>
<td>18.05137</td>
</tr>
<tr>
<td>26</td>
<td>17.009185</td>
<td>351.668352</td>
<td>214.58385</td>
<td>2.513282</td>
<td>4.260083</td>
<td>19.36172</td>
</tr>
<tr>
<td>28</td>
<td>23.293301</td>
<td>187.338921</td>
<td>137.30990</td>
<td>3.753809</td>
<td>2.936190</td>
<td>18.05137</td>
</tr>
<tr>
<td>29</td>
<td>24.776262</td>
<td>138.13614</td>
<td>118.17383</td>
<td>4.371926</td>
<td>1.426564</td>
<td>18.26750</td>
</tr>
<tr>
<td>30</td>
<td>22.308028</td>
<td>5.969799</td>
<td>118.17383</td>
<td>4.371926</td>
<td>1.426564</td>
<td>18.26750</td>
</tr>
<tr>
<td>31</td>
<td>19.228647</td>
<td>229.179303</td>
<td>148.90931</td>
<td>3.141172</td>
<td>3.222483</td>
<td>18.05137</td>
</tr>
<tr>
<td>32</td>
<td>25.065331</td>
<td>192.011334</td>
<td>142.97339</td>
<td>3.878321</td>
<td>2.142523</td>
<td>17.52720</td>
</tr>
<tr>
<td>33</td>
<td>26.419988</td>
<td>72.476092</td>
<td>79.17064</td>
<td>3.985763</td>
<td>2.283253</td>
<td>18.05137</td>
</tr>
<tr>
<td>34</td>
<td>9.768450</td>
<td>360.315682</td>
<td>190.86103</td>
<td>2.513282</td>
<td>4.260083</td>
<td>19.36172</td>
</tr>
<tr>
<td>35</td>
<td>17.009185</td>
<td>351.668352</td>
<td>214.58385</td>
<td>2.513282</td>
<td>4.260083</td>
<td>19.36172</td>
</tr>
<tr>
<td>37</td>
<td>23.293301</td>
<td>187.338921</td>
<td>137.30990</td>
<td>3.753809</td>
<td>2.936190</td>
<td>18.05137</td>
</tr>
<tr>
<td>38</td>
<td>24.776262</td>
<td>138.13614</td>
<td>118.17383</td>
<td>4.371926</td>
<td>1.426564</td>
<td>18.26750</td>
</tr>
<tr>
<td>39</td>
<td>22.308028</td>
<td>5.969799</td>
<td>118.17383</td>
<td>4.371926</td>
<td>1.426564</td>
<td>18.26750</td>
</tr>
<tr>
<td>X1</td>
<td>X2</td>
<td>X3</td>
<td>X4</td>
<td>X5</td>
<td>X6</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>16.624258</td>
<td>382.803044</td>
<td>192.96049</td>
<td>2.737288</td>
<td>4.5662243</td>
<td>15.58820</td>
<td></td>
</tr>
<tr>
<td>16.066352</td>
<td>306.918324</td>
<td>206.62547</td>
<td>3.236338</td>
<td>3.522393</td>
<td>15.67147</td>
<td></td>
</tr>
<tr>
<td>21.911348</td>
<td>211.557889</td>
<td>203.76420</td>
<td>3.814828</td>
<td>3.4294919</td>
<td>17.73663</td>
<td></td>
</tr>
<tr>
<td>23.432039</td>
<td>147.405877</td>
<td>62.34764</td>
<td>3.093330</td>
<td>4.1975334</td>
<td>17.59286</td>
<td></td>
</tr>
<tr>
<td>8.573387</td>
<td>424.640718</td>
<td>205.03377</td>
<td>3.087174</td>
<td>4.9942198</td>
<td>18.54980</td>
<td></td>
</tr>
<tr>
<td>16.592986</td>
<td>232.340174</td>
<td>119.36525</td>
<td>2.995524</td>
<td>4.2073236</td>
<td>18.54980</td>
<td></td>
</tr>
<tr>
<td>28.562558</td>
<td>136.617252</td>
<td>90.23244</td>
<td>4.337025</td>
<td>2.1406164</td>
<td>17.13835</td>
<td></td>
</tr>
<tr>
<td>27.511746</td>
<td>165.838291</td>
<td>67.28764</td>
<td>3.941838</td>
<td>1.6319913</td>
<td>16.21451</td>
<td></td>
</tr>
<tr>
<td>24.109918</td>
<td>202.000099</td>
<td>193.57244</td>
<td>3.169412</td>
<td>16.63171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.224393</td>
<td>439.424883</td>
<td>202.70133</td>
<td>3.251049</td>
<td>3.9443693</td>
<td>17.80867</td>
<td></td>
</tr>
<tr>
<td>25.981988</td>
<td>200.585319</td>
<td>138.01604</td>
<td>3.728431</td>
<td>2.8423040</td>
<td>19.31789</td>
<td></td>
</tr>
<tr>
<td>27.511746</td>
<td>219.025422</td>
<td>209.76230</td>
<td>3.814828</td>
<td>3.4294919</td>
<td>16.88691</td>
<td></td>
</tr>
<tr>
<td>24.109918</td>
<td>202.000099</td>
<td>193.57244</td>
<td>3.169412</td>
<td>16.63171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.224393</td>
<td>439.424883</td>
<td>202.70133</td>
<td>3.251049</td>
<td>3.9443693</td>
<td>17.80867</td>
<td></td>
</tr>
<tr>
<td>25.981988</td>
<td>200.585319</td>
<td>138.01604</td>
<td>3.728431</td>
<td>2.8423040</td>
<td>19.31789</td>
<td></td>
</tr>
</tbody>
</table>
7.2 Set of Nodes conditioning the node $X_1$

$h(X_1) = 3.2115821$ and $h(X_1|X_2, X_3, X_4, X_5, X_6) = 2.352636$

$h(X_1|X_2) = 2.572136$

$h(X_1|X_3) = 2.764256$

$h(X_1|X_4) = 2.903104$

$h(X_1|X_5) = 2.488241$

$h(X_1|X_6) = 3.186632$

The smallest conditional entropy is $h(X_1|X_3)$, we can compute the conditioning percentage:

$$
\lambda_{X_1|X_3} = \frac{h(X_1) - h(X_1|X_3)}{h(X_1)} = \frac{3.2115821 - 2.488241}{3.2115821 - 2.352636} = 0.8421263
$$

$h(X_1|X_3, X_2) = 2.435108$

$h(X_1|X_3, X_1) = 2.366194$

$h(X_1|X_3, X_4) = 2.487352$

$h(X_1|X_3, X_6) = 2.455447$

The smallest conditional entropy is $h(X_1|X_3, X_3)$, we can compute the conditioning percentage:

$$
\lambda_{X_1|X_3, X_3} = \frac{h(X_1) - h(X_1|X_3, X_3)}{h(X_1)} = \frac{3.2115821 - 2.366194}{3.2115821 - 2.352636} = 0.9842155
$$

$h(X_1|X_3, X_3, X_1) = 2.361073$

$h(X_1|X_3, X_3, X_4) = 2.365433$

$h(X_1|X_3, X_3, X_6) = 2.359072$

The smallest conditional entropy is $h(X_1|X_3, X_3, X_6)$, we can compute the conditioning percentage:

$$
\lambda_{X_1|X_3, X_3, X_6} = \frac{h(X_1) - h(X_1|X_3, X_3, X_6)}{h(X_1)} = \frac{3.2115821 - 2.353072}{3.2115821 - 2.352636} = 0.9925071
$$

15
\[ h(X_1|X_5, X_3, X_6, X_2) = 2.353589 \]
\[ h(X_1|X_5, X_3, X_6, X_4) = 2.358727 \]

The smallest conditional entropy is \( h(X_1|X_5, X_3, X_6, X_2) \), we can compute the conditioning percentage:

\[ \lambda_{X_1|X_5, X_3, X_6, X_2} = \frac{h(X_1) - h(X_1|X_5, X_3, X_6, X_2)}{h(X_1) - h(X_1|X_2, X_1, X_3, X_4, X_5, X_6)} = \frac{3.2115821 - 2.353589}{3.2115821 - 2.352636} = 0.9925071 \]

We can graph the conditioning percentage as a function of the conditioning:

![Conditioning percentages](image-url)
7.3 Set of nodes conditioning the node $X_2$

$h(X_2) = 6.275574$ and $h(X_2|X_1, X_3, X_4, X_5, X_6) = 5.30214$

$h(X_2|X_1) = 5.631888$

$h(X_2|X_3) = 5.787862$

$h(X_2|X_4) = 6.007003$

$h(X_2|X_5) = 5.477484$

$h(X_2|X_6) = 6.232157$

The smallest conditional entropy is $h(X_2|X_5)$, we can compute the conditioning percentage:

$$
\lambda_{X_2|X_5} = \frac{h(X_2) - h(X_2|X_5)}{h(X_2) - h(X_2|X_1, X_3, X_4, X_5, X_6)} = \frac{6.275574 - 5.477484}{6.275574 - 5.30214} = 0.8198707
$$

$h(X_2|X_5, X_1) = 5.424351$

$h(X_2|X_5, X_3) = 5.321245$

$h(X_2|X_5, X_4) = 5.46916$

$h(X_2|X_5, X_6) = 5.400476$

The smallest conditional entropy is $h(X_2|X_5, X_3)$, we can compute the conditioning percentage:

$$
\lambda_{X_2|X_5, X_3} = \frac{h(X_2) - h(X_2|X_5, X_3)}{h(X_2) - h(X_2|X_1, X_3, X_4, X_5, X_6)} = \frac{6.275574 - 5.321245}{6.275574 - 5.30214} = 0.9803736
$$

$h(X_2|X_5, X_3, X_1) = 5.316124$

$h(X_2|X_5, X_3, X_4) = 5.308246$

$h(X_2|X_5, X_3, X_6) = 5.320983$

The smallest conditional entropy is $h(X_2|X_5, X_3, X_4)$, we can compute the conditioning percentage:

$$
\lambda_{X_2|X_5, X_3, X_4} = \frac{h(X_2) - h(X_2|X_5, X_3, X_4)}{h(X_2) - h(X_2|X_1, X_3, X_4, X_5, X_6)} = \frac{6.275574 - 5.308246}{6.275574 - 5.30214} = 0.9937274
$$
The smallest conditional entropy is \( h(X_2|X_5, X_3, X_4, X_1) \), we can compute the conditioning percentage:

\[
\lambda_{X_2|X_5, X_3, X_4, X_1} = \frac{h(X_2) - h(X_2|X_5, X_3, X_4, X_6)}{h(X_2) - h(X_2|X_1, X_3, X_4, X_6)} = \frac{6.275574 - 5.308231}{6.275574 - 5.30214} = 0.99983305
\]

We can graph the conditioning percentage as a function of the conditioning:

![Figure 2: Conditioning percentages](image-url)
7.4 Set of nodes conditioning the node $X_3$

$h(X_3) = 5.555501$ and $h(X_3|X_1, X_2, X_4, X_5, X_6) = 4.726918$

$h(X_3|X_1) = 5.103935$

$h(X_3|X_2) = 5.067788$

$h(X_3|X_4) = 5.383439$

$h(X_3|X_5) = 5.22358$

$h(X_3|X_6) = 5.338916$

The smallest conditional entropy is $h(X_3|X_2)$, we can compute the conditioning percentage:

$$
\lambda_{X_3|X_2} = \frac{h(X_3) - h(X_3|X_2)}{h(X_3) - h(X_3|X_1, X_2, X_4, X_5, X_6)} = \frac{5.555501 - 5.067788}{5.555501 - 4.726918} = 0.5886109
$$

$h(X_3|X_2, X_1) = 5.017716$

$h(X_3|X_2, X_4) = 5.065635$

$h(X_3|X_2, X_5) = 5.067341$

$h(X_3|X_2, X_6) = 4.823889$

The smallest conditional entropy is $h(X_3|X_2, X_6)$, we can compute the conditioning percentage:

$$
\lambda_{X_3|X_2, X_6} = \frac{h(X_3) - h(X_3|X_2, X_6)}{h(X_3) - h(X_3|X_1, X_2, X_4, X_5, X_6)} = \frac{5.555501 - 4.823889}{5.555501 - 4.726918} = 0.8829677
$$

$h(X_3|X_2, X_6, X_1) = 4.734095$

$h(X_3|X_2, X_6, X_4) = 4.798265$

$h(X_3|X_2, X_6, X_5) = 4.804124$

The smallest conditional entropy is $h(X_3|X_2, X_6, X_1)$, we can compute the conditioning percentage:

$$
\lambda_{X_3|X_2, X_6, X_1} = \frac{h(X_3) - h(X_3|X_2, X_6, X_1)}{h(X_3) - h(X_3|X_1, X_2, X_4, X_5, X_6)} = \frac{5.555501 - 4.734095}{5.555501 - 4.726918} = 0.9913382
$$
$h(X_3|X_2, X_6, X_1, X_4) = 4.728869$

$h(X_3|X_2, X_6, X_1, X_5) = 4.734035$

The smallest conditional entropy is $h(X_3|X_2, X_6, X_1, X_4)$, we can compute the conditioning percentage:

$$\lambda_{X_3|X_2,X_6,X_1,X_4} = \frac{h(X_3) - h(X_3|X_2, X_6, X_1, X_4)}{h(X_3) - h(X_3|X_1, X_2, X_4, X_5, X_6)} = \frac{5.555501 - 4.728867}{5.555501 - 4.726918} = 0.9976478$$

We can graph the conditioning percentage as a function of the conditioning:

Figure 3: Conditioning percentages
7.5 Set of nodes conditioning the node $X_4$

$h(X_4) = 0.7993639$ and $h(X_4|X_1, X_2, X_3, X_5, X_6) = 0.3417083$

$h(X_4|X_1) = 0.486469$

$h(X_4|X_2) = 0.5307928$

$h(X_4|X_3) = 0.6273017$

$h(X_4|X_3) = 0.3616165$

$h(X_4|X_6) = 0.7975667$

The smallest conditional entropy is $h(X_4|X_5)$, we can compute the conditioning percentage:

$$\lambda_{X_1|X_5} = \frac{h(X_4) - h(X_4|X_5)}{h(X_4)} = \frac{0.7993639 - 0.3616165}{0.7993639 - 0.3417083} = 0.9564996$$

$h(X_4|X_5, X_1) = 0.3607268$

$h(X_4|X_5, X_2) = 0.3532928$

$h(X_4|X_5, X_3) = 0.3614813$

$h(X_4|X_5, X_6) = 0.3593926$

The smallest conditional entropy is $h(X_4|X_5, X_2)$, we can compute the conditioning percentage:

$$\lambda_{X_1|X_5, X_2} = \frac{h(X_4) - h(X_4|X_5, X_2)}{h(X_4)} = \frac{0.7993639 - 0.3532928}{0.7993639 - 0.3417083} = 0.9746873$$

$h(X_4|X_5, X_2, X_1) = 0.3493863$

$h(X_4|X_5, X_2, X_3) = 0.348482$

$h(X_4|X_5, X_2, X_6) = 0.3530991$

The smallest conditional entropy is $h(X_4|X_5, X_2, X_3)$, we can compute the conditioning percentage:

$$\lambda_{X_1|X_5, X_2, X_3} = \frac{h(X_4) - h(X_4|X_5, X_2, X_3)}{h(X_4)} = \frac{0.7993639 - 0.348482}{0.7993639 - 0.3417083} = 0.9851991$$

21
\[ h(X_4|X_5, X_2, X_1, X_1) = 0.3469016 \]
\[ h(X_4|X_5, X_3, X_6) = 0.3426615 \]

The smallest conditional entropy is \( h(X_4|X_5, X_2, X_3, X_6) \), we can compute the conditioning percentage:

\[ \lambda_{X_4|X_5,X_2,X_3,X_6} = \frac{h(X_4) - h(X_4|X_5, X_2, X_3, X_6)}{h(X_4) - h(X_4|X_1, X_2, X_3, X_5, X_6)} = \frac{0.7993639 - 0.3426615}{0.7993639 - 0.3417083} = 0.9979172 \]

We can graph the conditioning percentage as a function of the conditioning:

![Figure 4: Conditioning percentages](image-url)
7.6 Set of nodes conditioning the node $X_5$

$h(X_5) = 1.404689$ and $h(X_5|X_1, X_2, X_3, X_4, X_6) = 0.306058$

$h(X_5|X_1) = 0.6771087$

$h(X_5|X_2) = 0.6065983$

$h(X_5|X_3) = 1.072768$

$h(X_5|X_4) = 0.9669411$

$h(X_5|X_6) = 1.395599$

The smallest conditional entropy is $h(X_5|X_2)$, we can compute the conditioning percentage:

$$\lambda_{X_5|X_2} = \frac{h(X_5) - h(X_5|X_2)}{h(X_5) - h(X_5|X_1, X_2, X_3, X_4, X_6)} = \frac{1.404689 - 0.6065983}{1.404689 - 0.306058} = 0.7264411$$

$h(X_5|X_2, X_1) = 0.469571$

$h(X_5|X_2, X_3) = 0.6061512$

$h(X_5|X_2, X_4) = 0.4290984$

$h(X_5|X_2, X_6) = 0.5639188$

The smallest conditional entropy is $h(X_5|X_2, X_4)$, we can compute the conditioning percentage:

$$\lambda_{X_5|X_2, X_4} = \frac{h(X_5) - h(X_5|X_2, X_4)}{h(X_5) - h(X_5|X_1, X_2, X_3, X_4, X_6)} = \frac{1.404689 - 0.4290984}{1.404689 - 0.306058} = 0.8880057$$

$h(X_5|X_2, X_4, X_1) = 0.346577$

$h(X_5|X_2, X_4, X_3) = 0.4259941$

$h(X_5|X_2, X_4, X_6) = 0.401268$

The smallest conditional entropy is $h(X_5|X_2, X_4, X_1)$, we can compute the conditioning percentage:

$$\lambda_{X_5|X_2, X_4, X_1} = \frac{h(X_5) - h(X_5|X_2, X_4, X_1)}{h(X_5) - h(X_5|X_1, X_2, X_3, X_4, X_6)} = \frac{1.404689 - 0.346577}{1.404689 - 0.306058} = 0.9631186$$
The smallest conditional entropy is $h(X_5|X_2, X_4, X_1, X_6)$, we can compute the conditioning percentage:

$$\lambda_{X_5|X_1, X_2, X_4} = \frac{h(X_5) - h(X_5|X_2, X_4, X_1, X_6)}{h(X_5) - h(X_5|X_1, X_2, X_3, X_4, X_6)} = \frac{1.404689 - 0.3080077}{1.404689 - 0.306058} = 0.9982253$$

We can graph the conditioning percentage as a function of the conditioning:

![Figure 5: Conditioning percentages](image)
7.7 Set of nodes conditioning the node $X_6$

\[ h(X_6) = 1.954484 \text{ and } h(X_6|X_1, X_2, X_3, X_4, X_5) = 1.592492 \]
\[ h(X_6|X_1) = 1.922595 \]
\[ h(X_6|X_2) = 1.911067 \]
\[ h(X_6|X_3) = 1.737899 \]
\[ h(X_6|X_4) = 1.952687 \]
\[ h(X_6|X_5) = 1.945395 \]

The smallest conditional entropy is $h(X_6|X_3)$, we can compute the conditioning percentage:

\[ \lambda_{X_6|X_3} = \frac{h(X_6) - h(X_6|X_3)}{h(X_6) - h(X_6|X_1, X_2, X_3, X_4, X_5)} = \frac{1.954484 - 1.737899}{1.954484 - 1.592492} = 0.5983143 \]
\[ h(X_6|X_5, X_3, X_1) = 1.637045 \]
\[ h(X_6|X_5, X_3, X_2) = 1.667168 \]
\[ h(X_6|X_5, X_3, X_4) = 1.658338 \]
\[ h(X_6|X_5, X_3, X_5) = 1.605432 \]

The smallest conditional entropy is $h(X_6|X_3, X_5)$, we can compute the conditioning percentage:

\[ \lambda_{X_6|X_3, X_5} = \frac{h(X_6) - h(X_6|X_3, X_5)}{h(X_6) - h(X_6|X_1, X_2, X_3, X_4, X_5)} = \frac{1.954484 - 1.605432}{1.954484 - 1.592492} = 0.9642534 \]
\[ h(X_6|X_5, X_3, X_1) = 1.59831 \]
\[ h(X_6|X_5, X_3, X_2) = 1.60517 \]
\[ h(X_6|X_5, X_3, X_4) = 1.599364 \]

The smallest conditional entropy is $h(X_6|X_3, X_5, X_1)$, we can compute the conditioning percentage:

\[ \lambda_{X_6|X_3, X_5, X_1} = \frac{h(X_6) - h(X_6|X_3, X_5, X_1)}{h(X_6) - h(X_6|X_1, X_2, X_3, X_4, X_5)} = \frac{1.954484 - 1.59831}{1.954484 - 1.592492} = 0.9839278 \]
\[ h(X_6|X_3, X_5, X_1, X_2) = 1.597686 \]
\[ h(X_6|X_3, X_5, X_1, X_4) = 1.592658 \]

The smallest conditional entropy is \( h(X_6|X_3, X_5, X_1, X_4) \), we can compute the conditioning percentage:

\[
\lambda_{X_6|X_3, X_1, X_4} = \frac{h(X_6) - h(X_6|X_3, X_5, X_1, X_4)}{h(X_6) - h(X_6|X_1, X_2, X_3, X_4, X_5)} = \frac{1.954484 - 1.592658}{1.954484 - 1.592492} = 0.9995414
\]

We can graph the conditioning percentage as a function of the conditioning:

![Figure 6: Conditioning percentages](image-url)
7.8 Directed dependency graph obtained from a continuous data matrix

We set a conditioning percentage value of 95% ($\lambda = 0.95$) to be exceeded to obtain a good dependencies model:

The dependencies model expressed by the directed dependency graph can be expressed as follows:

$X_1 \not\perp\!\!\!\!\perp \{X_2, X_3\}$  \hspace{1cm} $\lambda_{X_1|x_2,x_3} = 0.9842155$

$X_2 \not\perp\!\!\!\!\perp \{X_3, X_5\}$  \hspace{1cm} $\lambda_{X_2|x_3,x_5} = 0.9803736$

$X_3 \not\perp\!\!\!\!\perp \{X_2, X_6, X_1\}$  \hspace{1cm} $\lambda_{X_3|x_2,x_6,x_1} = 0.9913382$

$X_4 \not\perp\!\!\!\!\perp \{X_3\}$  \hspace{1cm} $\lambda_{X_4|x_3} = 0.9564996$

$X_5 \not\perp\!\!\!\!\perp \{X_2, X_4, X_1\}$  \hspace{1cm} $\lambda_{X_5|x_2,x_4,x_1} = 0.9631186$

$X_6 \not\perp\!\!\!\!\perp \{X_3, X_5\}$  \hspace{1cm} $\lambda_{X_6|x_3,x_5} = 0.9642534$

where the symbol $\not\perp\!\!\!\!\perp$ corresponds to the dependency symbol.
8 Conclusion

We proposed a learning method for a directed dependency graph applied to a continuous data matrix. This method allowed us to extract the hidden directed dependencies in a continuous data matrix.

From a example, we explained step by step how to lead to the directed dependency graph.
References

