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 $\vec{x}$, defined on set $\mathbb{X}$ of size n, with a multidimensional probability density function (pdf) $p_X(\vec{x})$, we define the differential entropy $h(X)$ as:*

$$h(X) = -\int_{\mathbb{X}} p_X(\vec{x}) \ln p_X(\vec{x}) \overrightarrow{dx}$$

2.2 Joint differential entropy of two random vectors

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

$$h(X_1, X_2) = -\int_{\mathbb{X}_1} \int_{\mathbb{X}_2} p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \ln p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \overrightarrow{dx_1} \overrightarrow{dx_2}$$

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

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

$$h(X_1|X_2) = -\int_{\mathbb{X}_1} \int_{\mathbb{X}_2} p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \ln p_{X_1|X_2}(\vec{x}_1, \vec{x}_2) \overrightarrow{dx_1} \overrightarrow{dx_2}$$

where we have:

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

$$P_{X_2}(\vec{x}_2) = \int_{\mathbb{X}_1} p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \overrightarrow{dx_1}$$

2.4 Joint differential entropy and conditional differential entropy

Given two concatenated random vectors $(\vec{x}_1, \vec{x}_2)$, defined on the sets $\mathbb{X}_1$ and $\mathbb{X}_2$ of sizes n et m respectively, with a multidimensional probability density function (pdf) $p_{X_1 X_2}(\vec{x}_1, \vec{x}_2)$, 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:

$$
\begin{aligned}
& h(X_1|X_2) \\
&= -\int_{\mathbb{X}_1}\int_{\mathbb{X}_2} p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \ln p_{X_1|X_2}(\vec{x}_1|\vec{x}_2) \overrightarrow{dx_1}\overrightarrow{dx_2} \\
&= -\int_{\mathbb{X}_1}\int_{\mathbb{X}_2} p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \ln \frac{p_{X_1X_2}(\vec{x}_1, \vec{x}_2)}{p_{X_2}(\vec{x}_2)} \overrightarrow{dx_1}\overrightarrow{dx_2} \\
&= -\int_{\mathbb{X}_1}\int_{\mathbb{X}_2} P_{X_1X_2}(\vec{x}_1, \vec{x}_2) \ln P_{X_1X_2}(\vec{x}_1, \vec{x}_2) \overrightarrow{dx_1}\overrightarrow{dx_2} + \int_{\mathbb{X}_2} (\int_{\mathbb{X}_1} p_{X_1X_2}(\vec{x}_1, \vec{x}_2) \overrightarrow{dx_1}) \ln P_{X_2}(\vec{x}_2) \overrightarrow{dx_2} \\
&= -\int_{\mathbb{X}_1}\int_{\mathbb{X}_2} P_{X_1X_2}(\vec{x}_1, \vec{x}_2) \ln P_{X_1X_2}(\vec{x}_1, \vec{x}_2) \overrightarrow{dx_1}\overrightarrow{dx_2} + \int_{\mathbb{X}_2} P_{X_2}(\vec{x}_2) \ln P_{X_2}(\vec{x}_2) \overrightarrow{dx_2} \\
&= h(X_1, X_2) - h(X_2)
\end{aligned}
$$

2.5 Joint and conditional gaussian multidimensional probability

Consider a partioned random vector $\vec{x} = (\vec{x}_1, \vec{x}_2)$ of size $n = k_1 + k_2$, where k_1 and k_2 are the sizes of vectors $\vec{x}_1$ and $\vec{x}_2$ respectively, with a multivariate Gaussian distribution $P_X(\vec{x})$ with a mean vector $\vec{\mu_X}$ and covariance matrix K_{X^2}:

$$P_X(\vec{x}) = \mathcal{N}(\mu_X, K_{X^2}) = (2\pi)^{-\frac{n}{2}}|K_{X^2}|^{-\frac{1}{2}} \exp\{-\frac{(\vec{x}-\vec{\mu_X})^t K_{X^2}^{-1}(\vec{x}-\vec{\mu_X})}{2}\}$$

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

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

For this, we must start first from the block matrix multiplication of the covariance matrix K and the precision matrix $W = K^{-1}$ and prove the following relation:

$W_{X_2^2} = K_{X_2^2}^{-1} + W_{X_2X_1}.W_{X_1^2}^{-1}.W_{X_1X_2}$.

Indeed:

$$K_{X^2}W_{X^2} = \begin{pmatrix} K_{X_1^2}W_{X_1^2} + K_{X_1X_2}W_{X_2X_1} & K_{X_1^2}W_{X_1X_2} + K_{X_1X_2}W_{X_2^2} \\ K_{X_2X_1}W_{X_1^2} + K_{X_2^2}W_{X_2X_1} & K_{X_2X_1}W_{X_1X_2} + K_{X_2^2}W_{X_2^2} \end{pmatrix} = \begin{pmatrix} I_{k_1,k_1} & 0 \\ 0 & I_{k_2,k_2} \end{pmatrix}$$

$$K_{X_2X_1}W_{X_1^2} + K_{X_2^2}W_{X_2X_1} = 0$$

$$K_{X_2^2}^{-1}K_{X_2X_1}W_{X_1^2} + W_{X_2X_1} = 0$$

$$K_{X_2^2}^{-1}K_{X_2X_1} = -W_{X_2X_1}W_{X_1^2}^{-1}$$

$$K_{X_2^2}W_{X_2^2} + K_{X_2X_1}W_{X_1X_2} = I_{k_2k_2}$$

$$W_{X_2^2} = K_{X_2^2}^{-1} - K_{X_2}^{-1}.K_{X_2X_1}.W_{X_1X_2}$$

Finaly, we obtain:

$$W_{X_2^2} = K_{X_2^2}^{-1} + W_{X_2X_1}.W_{X_1^2}^{-1}.W_{X_1X_2}$$

Now, we will develop the Mahalanobis distance:

$$(\vec{x}-\vec{\mu_X})^t W_{X^2}(\vec{x}-\vec{\mu_X})$$

$$= (\vec{x}_1 - \vec{\mu_{X_1}}, \vec{x}_2 - \vec{\mu_{X_2}}) \begin{pmatrix} W_{X_1^2} & W_{X_1X_2} \\ W_{X_2X_1} & W_{X_2^2} \end{pmatrix} \begin{pmatrix} \vec{x}_1 - \vec{\mu_{X_1}} \\ \vec{x}_2 - \vec{\mu_{X_2}} \end{pmatrix}$$

$$= (\vec{x}_1 - \vec{\mu_{X_1}})^t W_{X_1^2}(\vec{x}_1 - \vec{\mu_{X_1}}) + (\vec{x}_1 - \vec{\mu_{X_1}})^t W_{X_1X_2}(\vec{x}_2 - \vec{\mu_{X_2}}) + (\vec{x}_2 - \vec{\mu_{X_2}})^t W_{X_2X_1}(\vec{x}_1 - \vec{\mu_{X_1}})$$

$$+(\vec{x}_2 - \vec{\mu_{X_2}})^t W_{X_2^2}(\vec{x}_2 - \vec{\mu_{X_2}})$$

Using the relation: $W_{X_2^2} = K_{X_2^2}^{-1} + W_{X_2X_1}.W_{X_1^2}^{-1}.W_{X_1X_2}$, we obtain:

$= (\vec{x_1} - \vec{\mu_{X_1}})^t W_{X_1^2} (\vec{x_1} - \vec{\mu_{X_1}}) + (\vec{x_1} - \vec{\mu_{X_1}})^t W_{X_1X_2} (\vec{x_2} - \vec{\mu_{X_2}}) + (\vec{x_2} - \vec{\mu_{X_2}})^t W_{X_2X_1} (\vec{x_1} - \vec{\mu_{X_1}})$

$+(\vec{x_2} - \vec{\mu_{X_2}})^t W_{X_2X_1} W_{X_1^2}^{-1} W_{X_1X_2} (\vec{x_2} - \vec{\mu_{X_2}}) + (\vec{x_2} - \vec{\mu_{X_2}})^t K_{X_2^2}^{-1} (\vec{x_2} - \vec{\mu_{X_2}})$

$= [(\vec{x_1} - \vec{\mu_{X_1}}) + W_{X_1^2}^{-1} W_{X_1X_2} (\vec{x_2} - \vec{\mu_{X_2}})]^t [W_{X_1^2} (\vec{x_1} - \vec{\mu_{X_1}}) + W_{X_1X_2} (\vec{x_2} - \vec{\mu_{X_2}})]$

$+(\vec{x_2} - \vec{\mu_{X_2}})^t K_{X_2^2}^{-1} (\vec{x_2} - \vec{\mu_{X_2}})$

$= [(\vec{x_1} - \vec{\mu_{X_1}}) + W_{X_1^2}^{-1} W_{X_1X_2} (\vec{x_2} - \vec{\mu_{X_2}})]^t W_{X_1^2}.[(\vec{x_1} - \vec{\mu_{X_1}}) + W_{X_1^2}^{-1} W_{X_1X_2} (\vec{x_2} - \vec{\mu_{X_2}})]$

$+(\vec{x_2} - \vec{\mu_{X_2}})^t K_{X_2^2}^{-1} (\vec{x_2} - \vec{\mu_{X_2}})$

We put:

$Q_1 = (\vec{x_1} - \vec{\nu_{X_1/X_2}})^t (K_{X_1^2} - K_{X_1X_2} K_{X_2^2}^{-1} K_{X_2X_1})^{-1} (\vec{x_1} - \vec{\nu_{X_1/X_2}})$

$\vec{\nu_{X_1|X_2}} = \vec{\mu_{X_1}} + K_{X_1X_2} K_{X_2^2}^{-1} (\vec{x_2} - \vec{\mu_{X_2}})$

$Q_2 = (\vec{x_2} - \vec{\mu_{X_2}})^t K_{X_2^2}^{-1} (\vec{x_2} - \vec{\mu_{X_2}})$

We then obtain the equalities as follows:

$(\vec{x} - \vec{\mu_X})^t K_{X^2}^{-1} (\vec{x} - \mu_X) = Q_1 + Q_2$

$P_X(\vec{x}) = (2\pi)^{-\frac{n}{2}} |K_{X^2}|^{-\frac{1}{2}} \exp\{-\frac{Q_1+Q_2}{2}\} = (2\pi)^{-\frac{n}{2}} |K_{X^2}|^{-\frac{1}{2}} \exp\{(\vec{x} - \vec{\mu_X})^t K_{X^2}^{-1} (\vec{x} - \vec{\mu_X})\}$

$P_{X_2}(\vec{x_2}) = (2\pi)^{-\frac{k_2}{2}} |K_{X_2^2}|^{-\frac{1}{2}} \exp\{-\frac{Q_2}{2}\} = (2\pi)^{-\frac{k_2}{2}} |K_{X_2^2}| \exp\{-\frac{(\vec{x_2} - \vec{\mu_{X_2}})^t K_{X_2^2}^{-1} (\vec{x_2} - \vec{\mu_{X_2}})}{2}\}$

Using the relation $\frac{P_{X_1X_2}(\vec{x_1}, \vec{x_2})}{P_{X_2}(x_2)}$:

$P_{X_1|X_2}(x_1, x_2) = (2\pi)^{-\frac{k_1}{2}} (\frac{|K_{(X_1X_2)^2}|}{|K_{X_2^2}|})^{-\frac{1}{2}} \exp\{-\frac{Q_1}{2}\} = (2\pi)^{-\frac{k_1}{2}} (\frac{|K_{(X_1X_2)^2}|}{|K_{X_2^2}|})^{-\frac{1}{2}} \exp\{-\frac{(\vec{x_1} - \vec{\nu_{X_1/X_2}})^t (K_{X_1^2} - K_{X_1X_2} K_{X_2^2}^{-1} K_{X_2X_1})^{-1} (\vec{x_1} - \vec{\nu_{X_1/X_2}})}{2}\}$

If we use the Schur's complement $K_{X_1^2|X_2} = K_{X_1^2} - K_{X_1X_2} K_{X_2^2}^{-1} K_{X_2X_1}$

we can express the conditional probability $P_{X_1|X_2}(\vec{x_1}, \vec{x_2})$ as follows:

$P_{X_1|X_2}(\vec{x_1}, \vec{x_2})$

$= (2\pi)^{-\frac{k_1}{2}} (\frac{|K_{(X_1X_2)^2}|}{|K_{X_2^2}|})^{-\frac{1}{2}} \exp\{-\frac{(\vec{x_1} - \vec{\nu_{X_1/X_2}})^t K_{X_1^2|X_2}^{-1} (\vec{x_1} - \vec{\nu_{X_1/X_2}})}{2}\} = (2\pi)^{-\frac{k_1}{2}} |K_{X_1^2|X_2}|^{-\frac{1}{2}} \exp\{-\frac{(\vec{x_1} - \vec{\nu_{X_1/X_2}})^t K_{X_1^2|X_2}^{-1} (\vec{x_1} - \vec{\nu_{X_1/X_2}})}{2}\}$

2.6 Differential entropy of a Gaussian random vector

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

$$P_X(\vec{x}) = \mathcal{N}(\mu_X, K_{X^2}) = (2\pi)^{-\frac{n}{2}}|K_{X^2}|^{-\frac{1}{2}} \exp\{-\frac{(\vec{x}-\vec{\mu_X})^t K_{X^2}^{-1}(\vec{x}-\vec{\mu_X})}{2}\}$$

with a mean vector μ_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:

$h(X)$

$= -\int_{-\infty}^{+\infty} p_X(\vec{x})\ln\{p_X(\vec{x})\}\overrightarrow{dx}$

$= -\int_{-\infty}^{+\infty} p_X(\vec{x})[-\frac{1}{2}(\vec{x}-\mu_X)^t K_{X^2}^{-1}(\vec{x}-\mu_X) - \ln(\sqrt{2\pi})^n|K_{X^2}|^{\frac{1}{2}}]\overrightarrow{dx}$

$= \frac{1}{2}E_X[\sum_{ij}(\vec{x}_i-\mu_{X_i})^t(K_{X^2}^{-1})_{ij}(\vec{x}_j-\mu_{X_j})] + \frac{1}{2}\ln(2\pi)^n|K_{X^2}|$

$= \frac{1}{2}E_X[\sum_{ij}(\vec{x}_i-\mu_{X_i})^t(\vec{x}_j-\mu_{X_j})(K_{X^2}^{-1})_{ij}] + \frac{1}{2}\ln(2\pi)^n|K_{X^2}|$

$= \frac{1}{2}\sum_{ij}E_X[(\vec{x}_j-\mu_{X_j})^t(\vec{x}_i-\mu_{X_i})](K_{X^2}^{-1})_{ij} + \frac{1}{2}\ln(2\pi)^n|K_{X^2}|$

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

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

$= \frac{1}{2}\sum_{j}I_{jj} + \frac{1}{2}\ln(2\pi)^n|K_{X^2}|$

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

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

2.7 Conditional differential entropy of two Gaussian random vectors

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

$$P_X(\vec{x}) = \mathcal{N}(\mu_X, K_{X^2}) = (2\pi)^{-\frac{n}{2}}|K_{X^2}|^{-\frac{1}{2}} \exp\{-\frac{(\vec{x} - \vec{\mu_X})^t K_{X^2}^{-1}(\vec{x} - \vec{\mu_X})}{2}\}$$

with a mean vector μ_X and a covariance matrix K_{X^2}.

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

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

Proof:

$h(X_1|X_2) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{X_1X_2}(\vec{x}_1, \vec{x}_2)\ln\{p_{X_1|X_2}(\vec{x}_1, \vec{x}_2)\}\overrightarrow{dx_1}\overrightarrow{dx_2}$

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

$P_{X_1|X_2}(\vec{x}_1, \vec{x}_2) = (2\pi)^{-\frac{k_1}{2}}|K_{X_1^2|X_2}|^{-\frac{1}{2}} \exp\{-\frac{(\vec{x}_1 - \vec{\nu_{X_1/X_2}})^t K_{X_1^2|X_2}^{-1}(\vec{x}_1 - \vec{\nu_{X_1/X_2}})}{2}\}$

So we can write:

$h(X_1|X_2)$

$= -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_{X_1X_2}(\vec{x}_1, \vec{x}_2)[-\frac{1}{2}(\vec{x}_1 - \nu_{X_1|X_2})^t K_{X^2}^{-1}(\vec{x}_1 - \nu_{X_1|X_2}) - \ln(\sqrt{2\pi})^{k_1}|K_{X_1^2|X_2}|^{\frac{1}{2}}]\overrightarrow{dx_1}\overrightarrow{dx_2}$

$= \frac{1}{2}E_{X_1X_2}[\sum_{ij}\{(\vec{x}_1)_i - \nu_{(X_1|X_2)_i}\}^t (K_{X_1^2|X_2}^{-1})_{ij}\{(\vec{x}_1)_j - \nu_{(X_1|X_2)_j}\}] + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

$= \frac{1}{2}E_{X_1X_2}[\sum_{ij}\{(\vec{x}_1)_i - \nu_{(X_1|X_2)_i}\}^t\{(\vec{x}_1)_j - \nu_{(X_1|X_2)_j}\}(K_{X_1^2|X_2}^{-1})_{ij}] + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

$= \frac{1}{2}\sum_{ij} E_{X_1X_2}[\{(\vec{x}_1)_j - \nu_{(X_1|X_2)_j}\}^t\{(\vec{x}_1)_i - \nu_{(X_1|X_2)_i}\}](K_{X_1^2|X_2}^{-1})_{ij} + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

$= \frac{1}{2}\sum_{ij}(K_{X_1^2|X_2})_{ji})(K_{X_1^2|X_2}^{-1})_{ij}) + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

$= \frac{1}{2}\sum_{j}(K_{X_1^2|X_2})_{jj})(K_{X_1^2|X_2}^{-1})_{jj}) + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

$= \frac{1}{2}\sum_{j} I_{jj} + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

$= \frac{k_1}{2} + \frac{1}{2}\ln(2\pi)^{k_1}|K_{X_1^2|X_2}|$

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

2.8 Inequalities theorem on the conditional differential entropies gaussian vectors

Theorem: *Given a partitioned Gaussian random vector $\vec{x} = (\vec{x}_1, \vec{x}_2, \vec{x}_3)$, of sizes k_1, k_2 and $k_3 = 1$ respectively, with the multivariate Gaussian distribution $\mathcal{N}(\mu_X, K_{X^2})$ 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 $K_{(X_1,X_2)^2}$ and the precision matrix $W_{(X_1,X_2)^2} = K^{-1}_{(X_1,X_2)^2}$ and prove the following relation:

$$W_{X_1^2} = K^{-1}_{X_1^2} + W_{X_1X_2}.W^{-1}_{X_2^2}.W_{X_2X_1}$$

$$K_{X^2}W_{X^2} = \begin{pmatrix} K_{X_1^2}W_{X_1^2} + K_{X_1X_2}W_{X_2X_1} & K_{X_1^2}W_{X_1X_2} + K_{X_1X_2}W_{X_2^2} \\ K_{X_2X_1}W_{X_1^2} + K_{X_2^2}W_{X_2X_1} & K_{X_2X_1}W_{X_1X_2} + K_{X_2^2}W_{X_2^2} \end{pmatrix} = \begin{pmatrix} I_{k_1,k_1} & 0 \\ 0 & I_{k_2,k_2} \end{pmatrix}$$

$K_{X_1X_2}.W_{X_2^2} + K_{X_1^2}.W_{X_1X_2} = 0$

$K^{-1}_{X_1^2}.K_{X_1X_2}.W_{X_2^2} + W_{X_1X_2} = 0$

$K^{-1}_{X_1^2}.K_{X_1X_2} = -W_{X_1X_2}.W^{-1}_{X_2^2}$

$K_{X_1^2}.W_{X_1^2} + K_{X_1X_2}.W_{X_2X_1} = I_{k_1k_1}$

$W_{X_1^2} = K^{-1}_{X_1^2} - K^{-1}_{X_1^2}.K_{X_1X_2}.W_{X_2X_1}$

$W_{X_1^2} = K^{-1}_{X_1^2} + W_{X_1X_2}.W^{-1}_{X_2^2}.W_{X_2X_1}$

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

$(K_{X_3,X_1}, K_{X_3X_2}).K^{-1}_{(X_1X_2)^2}.\begin{pmatrix} K_{X_1X_3} \\ K_{X_2X_3} \end{pmatrix}$

$= (K_{X_3,X_1}, K_{X_3X_2}).W_{(X_1X_2)^2}.\begin{pmatrix} K_{X_1X_3} \\ K_{X_2X_3} \end{pmatrix}$

$= (K_{X_3X_1}, K_{X_3X_2})\begin{pmatrix} W_{X_1^2} & W_{X_1X_2} \\ W_{X_2X_1} & W_{X_2^2} \end{pmatrix}\begin{pmatrix} K_{X_1X_3} \\ K_{X_2X_3} \end{pmatrix}$

$= (K_{X_3X_1})^t W_{X_1^2}(K_{X_1X_3}) + (K_{X_3X_1})W_{X_1X_2}(K_{X_2X_3}) + (K_{X_3X_2})W_{X_2X_1}(K_{X_1X_3})$

$+(K_{X_3X_2}).W_{X_2^2}(K_{X_2X_3})$

Using the relation: $W_{X_1^2} = K^{-1}_{X_1^2} + W_{X_1X_2}.W^{-1}_{X_2^2}.W_{X_2X_1}$:

$= (K_{X_3X_1}).K^{-1}_{X_1^2}.(K_{X_1X_3}) + (K_{X_3X_1}).W_{X_1X_2}.W^{-1}_{X_2^2}.W_{X_2X_1}.(K_{X_1X_3})$

$+(K_{X_3X_1}).W_{X_1X_2}.(K_{X_2X_3}) + (K_{X_3X_2}).W_{X_2X_1}.(K_{X_1X_3})$

$+(K_{X_3X_2}).W_{X_2^2}.(K_{X_2X_3})$

$= [(K_{X_2X_3}) + W_{X_2^2}^{-1}.W_{X_2X_1}.(K_{X_1X_3})]^t.[W_{X_2^2}.(K_{X_2X_3}) + W_{X_2X_1}.(K_{X_1X_3})]$

$+(K_{X_3X_1}).K_{X_1^2}^{-1}.(K_{X_1X_3})$

$= (K_{X_3X_1}) \cdot K_{X_1^2}^{-1}.(K_{X_1X_3})$

$+[(K_{X_2X_3}) + W_{X_2^2}^{-1}.W_{X_2X_1}.(K_{X_1X_3})]^t.W_{X_2^2}.[(K_{X_2X_3}) + W_{X_2^2}^{-1}.W_{X_2X_1}.(K_{X_1X_3})]$

However both following quadratics forms are equivalents :

$$(K_{X_3X_1}, K_{X_3X_2})\begin{pmatrix} W_{X_1^2} & W_{X_1X_2} \\ W_{X_2X_1} & W_{X_2^2} \end{pmatrix}\begin{pmatrix} K_{X_1X_3} \\ K_{X_2X_3} \end{pmatrix} = (K_{X_3X_2}, K_{X_3X_1})\begin{pmatrix} W_{X_2^2} & W_{X_2X_1} \\ W_{X_1X_2} & W_{X_1^2} \end{pmatrix}\begin{pmatrix} K_{X_2X_3} \\ K_{X_1X_3} \end{pmatrix}$$

and are positive semidefinite if and only if:

$W_{X_1^2} \geq 0$, $W_{X_2^2} - W_{X_2X_1}.W_{X_1^2}^{-1}.W_{X_1X_2} \geq 0$

but yet

$W_{X_2^2} \geq 0$, $W_{X_1^2} - W_{X_1X_2}.W_{X_2^2}^{-1}.W_{X_2X_1} \geq 0$

As $W_{X_2^2} \geq 0$, we can write the inequalities:

$$(K_{X_3,X_1}, K_{X_3X_2}) \cdot K_{(X_1X_2)^2}^{-1} \cdot \begin{pmatrix} K_{X_1X_3} \\ K_{X_2X_3} \end{pmatrix} \geq (K_{X_3X_1}).K_{X_1^2}^{-1}.(K_{X_1X_3}) \geq 0$$

If we use $K_{X_3^2}$, we can write:

$$K_{X_3^2} - (K_{X_3,X_1}, K_{X_3X_2}).K_{(X_1X_2)^2}^{-1} \cdot \begin{pmatrix} K_{X_1X_3} \\ K_{X_2X_3} \end{pmatrix} \leq K_{X_3^2} - (K_{X_3X_1}).K_{X_1^2}^{-1}.(K_{X_1X_3}) \leq K_{X_3^2}$$

$$K_{X_3^2|X_1X_2} \leq K_{X_3^2|X_1} \leq K_{X_3^2}$$

$$\frac{1}{2}\ln|K_{X_3^2|X_1X_2}| + \frac{1}{2}\ln(2\pi e)^n \leq \frac{1}{2}\ln|K_{X_3^2|X_1}| + \frac{1}{2}\ln(2\pi e)^n \leq \frac{1}{2}\ln|K_{X_3^2}| + \frac{1}{2}\ln(2\pi e)^n$$

Finally, we have the relation:

$$h(X_3|X_1, X_2) \leq h(X_3|X_1) \leq h(X_3)$$

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 $\Omega \equiv X_{i=1,\dots,n}$, the variable $X_j \in \Omega$, the subsets $\Omega_1 \subset \Omega \backslash X_j$ and $\Omega_2 \equiv \Omega \backslash \{X_j, \Omega_1\}$ we can define the conditioning percentage $\lambda_{X_j|\Omega_1}$ of the variables Ω_1 which act on the variable X_j as follows:*

$$\lambda_{X_j|\Omega_1} = \frac{h(X_j) - h(X_j|\Omega_1)}{h(X_j) - h(X_j|\Omega_1,\Omega_2)} = \frac{E_\Omega[\ln P_{X_j|\Omega_1}(x_j, \vec{\omega}_1)]}{h(X_j) - h(X_j|\Omega_1,\Omega_2)} + \frac{h(X_j)}{h(X_j) - h(X_j|\Omega_1,\Omega_2)}$$

$$0 \leq \lambda_{X_j|\Omega_1} \leq 1$$

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 $h(X_j|\Omega_1,\Omega_2) = h(X_j|\Omega_1) = h(X_j)$. 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 $\Omega \equiv X_{i=1,\dots,n}$, the Gaussian variable $X_j \in \Omega$, the subsets $\Omega_1 \subset \Omega \backslash X_j$ and $\Omega_2 \equiv \Omega \backslash \{X_j, \Omega_1\}$ we can define the Gaussian conditioning percentage $\lambda_{X_j|\Omega_1}$ of the Gaussian variables Ω_1 which act on the Gaussian variable X_j as follows:*

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

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 Ω_1 *to the node* X_j*:*

$$\lambda_{X_j|\Omega_1} = \frac{h(X_j) - h(X_j|\Omega_1)}{h(X_j) - h(X_j|\Omega_1, \Omega_2)} = \frac{E_\Omega[\ln P_{X_j|\Omega_1}(x_j, \vec{\omega}_1)]}{h(X_j) - h(X_j|\Omega_1, \Omega_2)} + \frac{h(X_j)}{h(X_j) - h(X_j|\Omega_1, \Omega_2)}$$

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:

$$\begin{array}{ll}
\max\limits_{X_j} \lambda_{X_i|X_j} & \min\limits_{X_j} h(X_i|X_j) \\
\max\limits_{X_k} \lambda_{X_i|X_j,X_k} & \min\limits_{X_k} h(X_i|X_j, X_k) \\
\max\limits_{X_l} \lambda_{X_i|X_j,X_k,X_l} & \min\limits_{X_l} h(X_i|X_j, X_k, X_l) \\
\vdots & \vdots
\end{array}$$

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

X1	X2	X3	X4	X5	X6
21.697356	212.496303	100.27983	4.067217	3.1128370	20.45330
17.933487	334.547171	216.25136	3.032607	3.9452276	17.51779
22.593178	293.789279	131.11323	3.847017	2.8745655	17.27431
34.049362	-140.459877	59.76671	5.185856	0.3781136	18.91340
18.893331	193.854070	165.98525	3.619441	2.8882926	17.09582
27.386443	183.449699	89.49365	4.120053	2.2747225	19.61326
29.387658	-27.047273	48.51673	4.005462	1.2374293	19.60590
13.803899	289.576913	203.10891	2.691737	4.4497561	17.63801
23.307997	190.350364	83.14425	4.433126	2.3252469	21.52021
34.096057	2.741246	47.94401	3.978904	0.8459643	16.59992
19.337734	229.179303	148.90931	3.141172	3.2224283	18.05137
12.740558	332.567200	198.15682	3.502937	3.5581512	17.08335
19.019523	177.643152	75.71239	3.984464	3.0206339	19.12922
14.515920	251.345140	238.63220	3.392142	3.8888359	15.05471
24.641912	156.073251	172.47024	3.922760	2.1692936	17.03247
22.308028	5.969799	118.17383	4.371926	1.4265646	18.26750
17.009185	351.668352	214.58385	2.698569	4.2832532	17.28782
19.228647	256.121885	158.85563	4.233123	3.0393819	17.76284
25.065331	192.011334	184.91772	3.628895	2.3155048	18.26881
26.441899	158.193829	115.22245	4.897830	2.2291997	17.05538
24.921998	72.476092	79.17064	3.897563	2.2206522	18.50567
9.768450	360.315682	190.86103	2.513282	4.2600833	19.36172
21.708682	230.343087	148.22627	3.872064	2.9246166	18.02496
23.293301	187.338921	137.30990	3.753809	2.9361905	16.60604
24.776262	102.153614	141.87334	3.991304	2.1954443	16.27751
17.292975	209.795454	103.01824	3.078183	3.4375820	18.70655
20.456419	148.953697	142.97339	3.878322	2.1425225	17.52720
20.389620	186.221825	172.58877	3.685490	2.6114183	17.69306
24.153918	88.997109	73.04434	3.855536	2.7443201	21.19175
12.366297	345.902899	223.52569	2.889015	3.6826036	16.73392
13.257675	499.108686	204.25494	2.968723	4.7720203	19.52203
22.299240	89.665515	122.32289	3.868410	2.5990239	18.06212
8.903630	459.562948	255.89935	3.511985	4.6287205	16.20319
16.803197	339.644711	156.85455	3.236518	3.9164802	16.86312
23.318325	218.850544	121.29319	3.717060	2.9066304	17.16336
19.983920	217.473115	149.56058	3.144869	3.5337949	17.99537
20.879636	175.264826	189.29793	3.727292	2.5445839	17.28957
13.007037	328.143139	188.78226	2.739336	4.3101301	18.32834
19.705524	255.280329	166.26701	3.927320	3.3045594	19.21580
12.909625	371.599198	208.89409	2.813123	4.3670589	20.09771
26.607515	-11.397382	90.50657	4.801016	1.7515093	18.51810
18.357485	273.627256	110.10703	3.721136	3.5930101	19.25385
21.734258	194.486630	147.25357	4.265773	2.7584258	15.87150
18.679990	166.121706	153.32705	3.813516	2.9196218	17.03728
18.013215	281.022416	170.99185	3.164209	3.6305637	19.03299
25.210935	156.325110	108.98786	4.197637	1.9370005	18.86877
25.474886	120.080876	86.99225	3.575362	3.1195229	19.01564
25.558855	149.638937	186.34757	4.247228	2.3641699	14.06583
19.093939	241.095004	123.48452	3.460038	3.6587681	18.50821
26.383655	144.062768	114.88925	4.624856	2.2478064	18.06069
16.631879	415.350156	202.13384	3.246128	4.3677528	18.03676
17.429027	499.253634	227.75786	3.007525	5.0938728	16.08398

X1	X2	X3	X4	X5	X6
16.624258	382.803044	192.96049	2.737288	4.5662243	15.58820
16.066352	306.918324	206.62547	3.236338	3.5222393	15.67147
17.504180	219.025422	209.76230	3.814828	3.4294919	16.88691
21.911348	211.557889	78.55344	4.055491	2.9154360	18.01431
26.577185	144.955804	113.40457	4.097286	2.5707701	17.95062
10.088415	358.589760	231.47066	3.259937	4.1975334	17.73663
23.432039	147.405877	62.37464	3.093330	3.4331099	21.53923
8.573387	424.640718	205.03377	3.087174	4.9942198	17.25097
16.592986	232.340174	119.36525	2.995524	4.2073326	18.54980
28.562558	136.617252	90.23244	4.337025	2.1406164	17.13835
27.511746	165.838291	67.28764	3.941838	1.6319913	16.21451
24.109918	202.000099	193.57244	4.109444	3.1691412	16.63171
15.224393	439.424883	202.70133	3.251049	3.9443693	17.80867
25.981988	200.585319	138.01604	3.728482	3.4324030	19.31789
22.375536	138.939767	120.75469	3.566029	3.3694704	18.77820
18.302730	312.679313	273.84440	3.401653	3.9555278	15.35012
16.496028	278.680945	126.55229	4.185533	3.1401651	18.77259
10.648459	330.025619	237.83779	3.005619	4.2221736	18.54193
19.577220	318.735572	176.17215	3.228952	3.3865012	17.64093
12.602861	547.610755	325.18760	3.364679	4.5170861	16.94013
23.483835	269.181617	221.92444	3.724709	2.8467341	13.64431
11.520018	436.855756	264.40488	3.735643	4.7439170	15.15520
22.735042	239.556194	103.32283	4.569178	2.6527874	17.80767
32.766632	84.141232	54.03472	4.122624	1.3440861	19.77657
3.368601	560.702187	276.03769	2.742008	5.5880158	16.24667
18.582352	223.744131	92.24361	3.101948	3.8218877	19.78210
25.259709	142.345346	102.57292	4.157116	2.5808681	18.65086
25.862437	91.731591	115.25040	3.725904	2.9508191	18.51111
24.405828	264.105590	128.24257	3.379417	3.7308016	17.89932
26.484348	198.879063	195.17488	3.908468	2.1831227	13.83249
27.577924	131.411099	88.60592	3.664482	2.3816511	19.36957
16.500382	320.080511	143.27535	3.873874	3.9361777	18.36889
19.695883	197.885483	122.45748	3.494311	2.8443474	16.61144
18.822765	336.902803	192.74357	3.289000	3.7248344	16.63973
23.940722	206.683049	157.44793	3.657017	3.1324352	19.14949
11.244597	445.089964	244.00301	3.483708	4.5950475	17.16837
12.142727	446.114532	198.29972	3.270059	4.5206776	16.62513
17.717054	262.381335	158.13466	3.228514	4.1851491	19.39476
27.938884	122.305509	23.85977	3.736659	3.0321913	22.76239
23.992895	88.638377	66.26240	4.009436	2.6843940	18.06440
18.545182	309.421665	150.63214	3.405085	3.3540838	19.26326
29.511690	189.962882	88.99168	4.235150	2.0802414	19.23400
24.473435	157.837949	158.95482	4.284887	2.1564811	13.37018
15.336709	307.162666	206.02291	3.002996	3.2221951	18.15282
36.603287	-52.933877	-54.60674	4.250180	1.3984221	23.06948
20.055056	238.642997	164.02440	3.947206	2.8032071	16.87895
21.648469	188.946992	138.74106	3.861113	2.8219705	18.50646
17.156453	353.615611	121.96272	4.664249	3.8069832	19.12094

7.2 Set of Nodes conditioning the node X_1

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

$h(X_1|X_2) = 2.572136$

$h(X_1|X_3) = 2.764256$

$h(X_1|X_4) = 2.903104$

$h(X_1|X_5) = 2.488241$

$h(X_1|X_6) = 3.186632$

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

$$\lambda_{X_1|X_5} = \frac{h(X_1) - h(X_1|X_5)}{h(X_1) - h(X_1|X_2, X_3, X_3, X_4, X_5, X_6)} = \frac{3.2115821 - 2.488241}{3.2115821 - 2.352636} = 0.8421263$$

$h(X_1|X_5, X_2) = 2.435108$

$h(X_1|X_5, X_3) = 2.366194$

$h(X_1|X_5, X_4 = 2.487352$

$h(X_1|X_5, X_6) = 2.455447$

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

$$\lambda_{X_1|X_5,X_3} = \frac{h(X_1) - h(X_1|X_5, X_3)}{h(X_1) - h(X_1|X_2, X_3, X_3, X_4, X_5, X_6)} = \frac{3.2115821 - 2.366194}{3.2115821 - 2.352636} = 0.9842155$$

$h(X_1|X_5, X_3, X_2) = 2.361073$

$h(X_1|X_5, X_3, X_4) = 2.365433$

$h(X_1|X_5, X_3, X_6) = 2.359072$

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

$$\lambda_{X_1|X_5,X_3,X_6} = \frac{h(X_1) - h(X_1|X_5, X_3, X_6)}{h(X_1) - h(X_1|X_2, X_3, X_3, X_4, X_5, X_6)} = \frac{3.2115821 - 2.353072}{3.2115821 - 2.352636} = 0.9925071$$

$h(X_1|X_5, X_3, X_6, X_2) = 2.353589$

$h(X_1|X_5, X_3, X_6, X_4) = 2.358727$

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

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

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

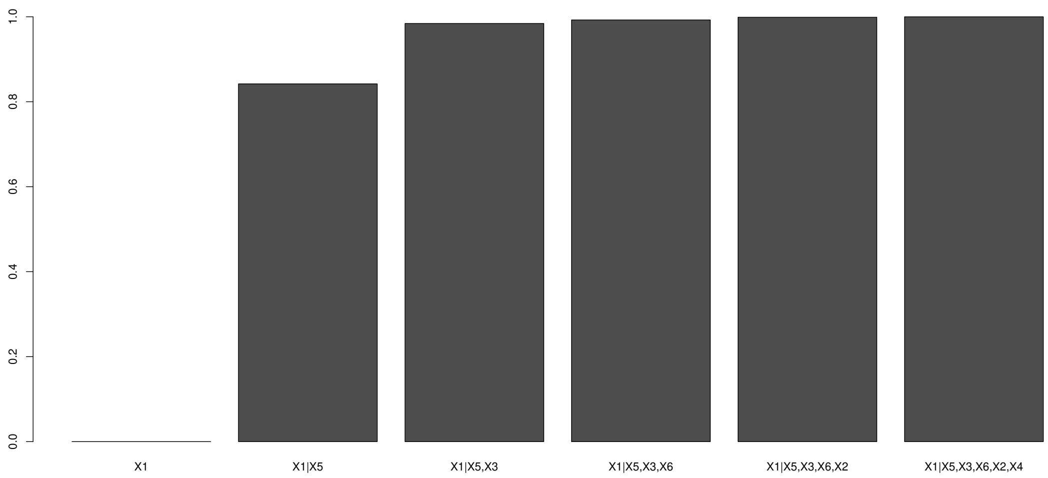


Figure 1: Conditioning percentages

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 :

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

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

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

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

$h(X_1|X_5, X_6) = 5.400476$

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 :

$$\lambda_{X_2|X_5,X_3,X_4} = \frac{h(X_2) - h(X_2|X_5, X_3, X_4)}{h(X_2) - h(X_2|X_1, X_3, X_4, X_5, X_6)} = \frac{6.275574 - 5.308246}{6.275574 - 5.30214} = 0.9937274$$

$h(X_2|X_5, X_3, X_4, X_1) = 5.302305$

$h(X_2|X_5, X_3, X_4, X_6) = 5.308231$

The smallest conditional entropy is $h(X_2|X_5, X_3, X_4, X_1)$, we can compute the conditioning percentage :

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

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

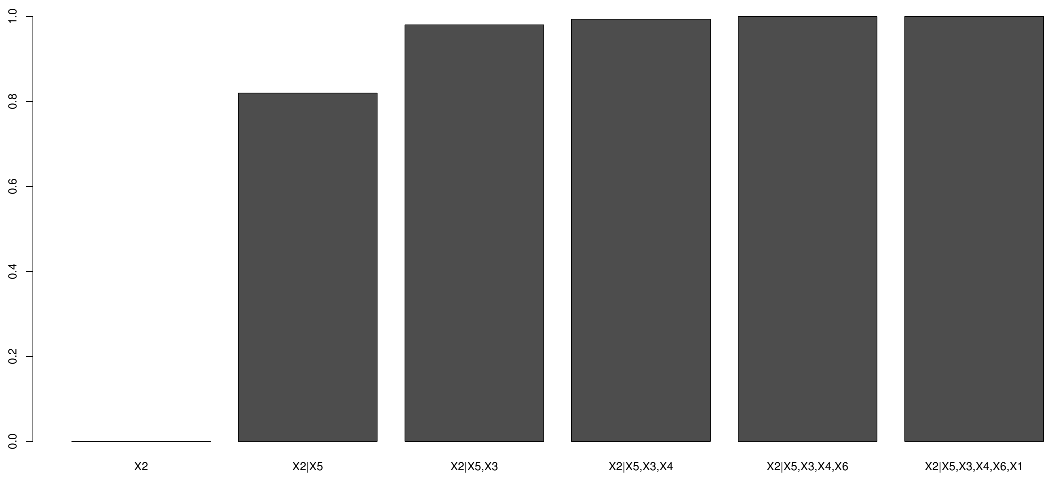


Figure 2: Conditioning percentages

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:

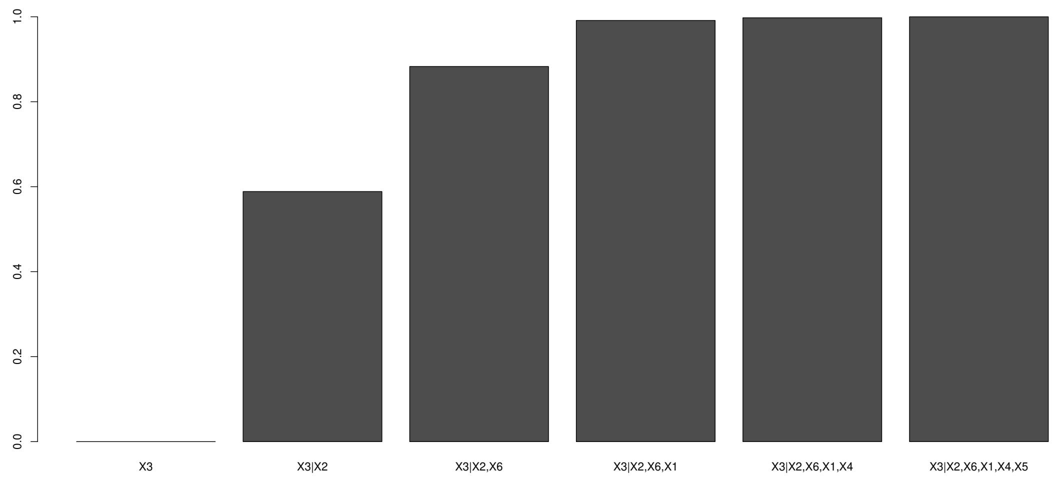


Figure 3: Conditioning percentages

7.5 Set of nodes conditioning the node X_4

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

$h(X_4|X_1) = 0.4866469$

$h(X_4|X_2) = 0.5307928$

$h(X_4|X_3) = 0.6273017$

$h(X_4|X_5) = 0.3616165$

$h(X_4|X_6) = 0.7975667$

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

$$\lambda_{X_4|X_5} = \frac{h(X_4) - h(X_4|X_5)}{h(X_4) - h(X_4|X_1, X_2, X_3, X_5, X_6)} = \frac{0.7993639 - 0.3616165}{0.7993639 - 0.3417083} = 0.9564996$$

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

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

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

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

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

$$\lambda_{X_4|X_5,X_2} = \frac{h(X_4) - h(X_4|X_5, X_2)}{h(X_4) - h(X_4|X_1, X_2, X_3, X_5, X_6)} = \frac{0.7993639 - 0.3532928}{0.7993639 - 0.3417083} = 0.9746873$$

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

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

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

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

$$\lambda_{X_4|X_5,X_2,X_3} = \frac{h(X_4) - h(X_4|X_5, X_2, X_3)}{h(X_4) - h(X_4|X_1, X_2, X_3, X_5, X_6)} = \frac{0.7993639 - 0.348482}{0.7993639 - 0.3417083} = 0.9851991$$

$h(X_4|X_5, X_2, X_3, X_1) = 0.3469016$

$h(X_4|X_5, X_2, X_3, X_6) = 0.3426615$

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:

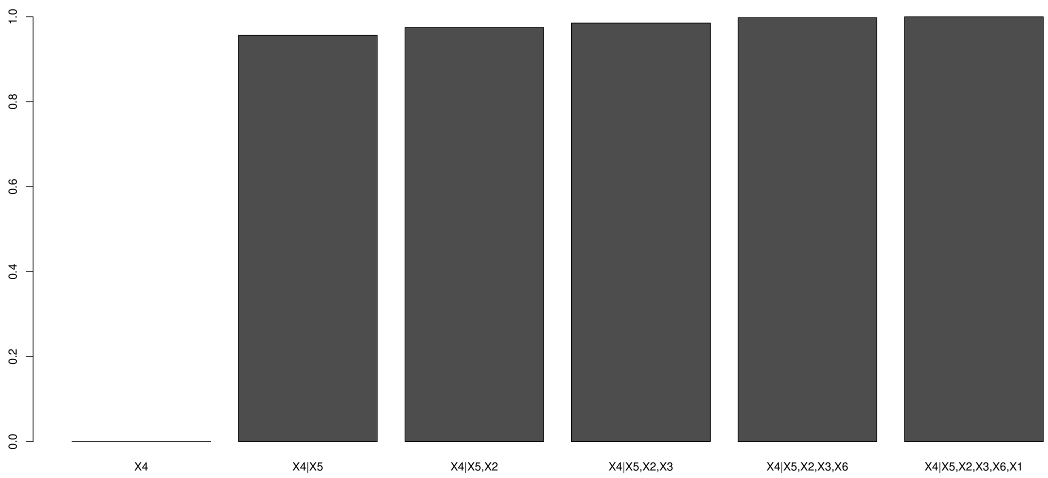


Figure 4: Conditioning percentages

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 :

$$\lambda_{X_5|X_2,X_4,X_1} = \frac{h(X_5) - h(X_5|X_2, X_4, X_1)}{h(X_5) - h(X_5|X_1, X_2, X_3, X_4, X_6)} = \frac{1.404689 - 0.346577}{1.404689 - 0.306058} = 0.9631186$$

$h(X_5|X_2, X_4, X_1, X_3) = 0.3205307$

$h(X_5|X_2, X_4, X_1, X_6) = 0.3080077$

The smallest conditional entropy is $h(X_5|X_2, X_4, X_1, X_6)$, we can compute the conditioning percentage :

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

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

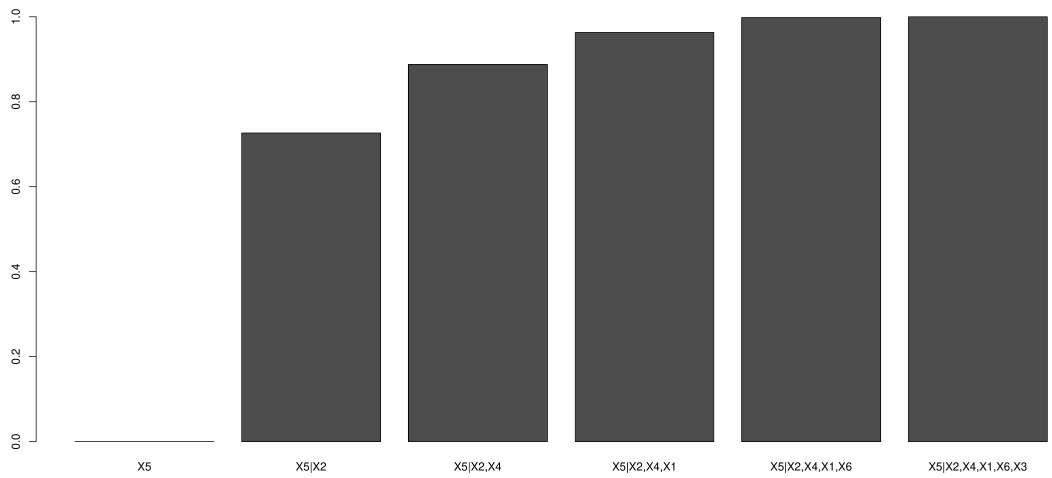


Figure 5: Conditioning percentages

7.7 Set of nodes conditioning the node X_6

$h(X_6) = 1.954484$ and $h(X_6|X_1, X_2, X_3, X_4, X_5) = 1.592492$

$h(X_6|X_1) = 1.9225295$

$h(X_6|X_2) = 1.911067$

$h(X_6|X_3) = 1.737899$

$h(X_6|X_4) = 1.952687$

$h(X_6|X_5) = 1.945395$

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

$$\lambda_{X_6|X_3} = \frac{h(X_6) - h(X_6|X_3)}{h(X_6) - h(X_6|X_1, X_2, X_3, X_4, X_5)} = \frac{1.954484 - 1.737899}{1.954484 - 1.592492} = 0.5983143$$

$h(X_6|X_3, X_1) = 1.637045$

$h(X_6|X_3, X_2) = 1.667168$

$h(X_6|X_3, X_4) = 1.658338$

$h(X_6|X_3, X_5) = 1.605432$

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

$$\lambda_{X_6|X_3,X_5} = \frac{h(X_6) - h(X_6|X_3, X_5)}{h(X_6) - h(X_6|X_1, X_2, X_3, X_4, X_5)} = \frac{1.954484 - 1.605432}{1.954484 - 1.592492} = 0.9642534$$

$h(X_6|X_3, X_5, X_1) = 1.59831$

$h(X_6|X_3, X_5, X_2) = 1.60517$

$h(X_6|X_3, X_5, X_4) = 1.599364$

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 :

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

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

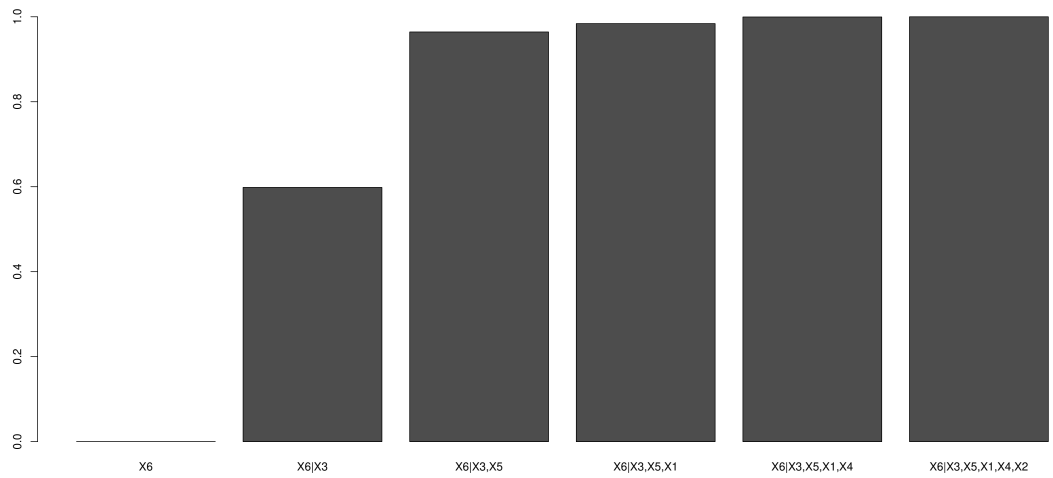
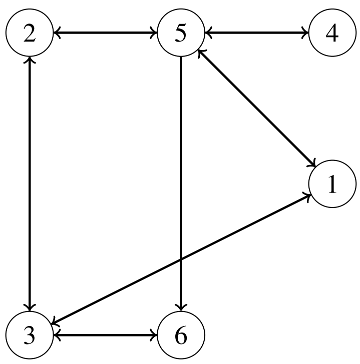


Figure 6: Conditioning percentages

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:

$X_1 \not\perp \{X_5, X_3\} \qquad \lambda_{X_1|X_5,X_3} = 0.9842155$

$X_2 \not\perp \{X_5, X_3\} \qquad \lambda_{X_2|X_5,X_3} = 0.9803736$

$X_3 \not\perp \{X_2, X_6, X_1\} \qquad \lambda_{X_3|X_2,X_6,X_1} = 0.9913382$

$X_4 \not\perp \{X_5\} \qquad \lambda_{X_3|X_5} = 0.9564996$

$X_5 \not\perp \{X_2, X_4, X_1\} \qquad \lambda_{X_5|X_2,X_4,X_1} = 0.9631186$

$X_6 \not\perp \{X_3, X_5\} \qquad \lambda_{X_6|X_3,X_5} = 0.9642534$

where the symbol $\not\perp$ 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

[1] Elements of information theory. Author: Thomas M.Cover and Joy A.Thomas. Copyright 1991 John Wiley and sons.

[2] Optimal stastical decisions. Author: Morris H.DeGroot. Copyright 1970-2004 John Wiley and sons.

[3] Matrix Analysis. Author: Roger A.Horn and Charles R.Johnson. Copyright 2012, Cambridge university press.