Master MSIAM M2

Fundamentals of Probabilistic Data Mining

Lab works: probabilistic graphical models

`http://chamilo2.grenet.fr/inp/courses/ENSIMAGWMM9MO17`

This laboratory work aims at comparing two procedures to estimate multivariate Gaussian directed probabilistic graphical models. It relies on the `R bnlearn` package for structural inference.

Run `R` and test whether the `bnlearn` package is installed by trying the command `library(bnlearn)`. If it fails, run `install.packages("bnlearn")`. Similarly, test whether the `Rgraphviz` package is installed by `library(Rgraphviz)`. If it fails, run `source("https://bioconductor.org/biocLite.R")` and then `biocLite("Rgraphviz")`.

1 Simulated data

Firstly, simulate a Gaussian model with the perfect map in Figure 1. To do this, use linear regression models using offsets 0, the coefficients in the figure, and the residual standard deviation $\sigma = 1$. Simulate one sample of size 40 and one sample of size 100.

Compare the `gs` and `hc` procedures (Scutari, 2010) using both sample sizes. What is your conclusion?

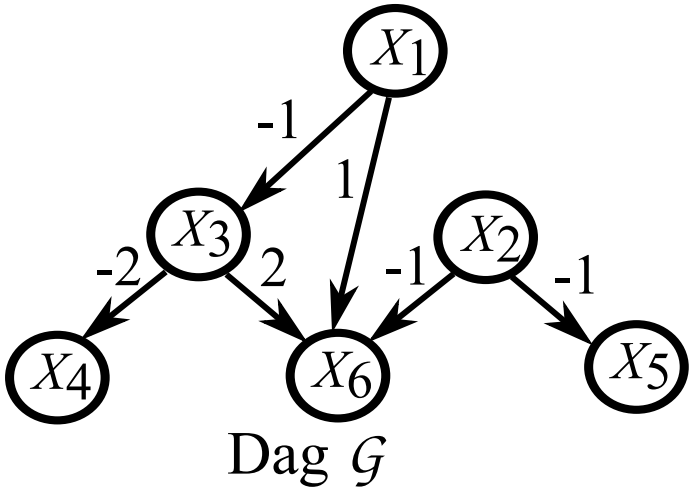

Figure 1: Simulated model

2 Real data: asset returns

We consider the returns of 8 assets on $n = 5,039$ days. The daily return $X_{t,i}$ of asset i at time t is defined as $(V_{t,i} - V_{t-1,i})/V_{t-1,i}$, where $V_{t,i}$ is the value of asset i at time t.

Here, we consider the assets "AIR.FRANCE.KLM", "ALCATEL.LUCENT" "AXA", "FAURECIA", "GAUMONT", "GEODIS", "PPR" and "UNION.FINC.FRANC." only.

1. Use file "Returns250d.txt" to create a data frame with only the 8 assets listed above.

2. Estimate directed graphs using the `gs` and `hc` procedures (Scutari, 2010) and plot their graphs.

3. Find a marginal independence relationship between two variables found by `gs` but not by `hc`. Use `ci.test` to perform a statistical test of independence. What do you conclude?

4. Find a conditional independence relationship between two variables given another set of variables found by `hc` but not by `gs`. Use `ci.test` to perform a statistical test of (conditional) independence. What do you conclude?

5. Find a conditional independence relationship between two variables given another set of variables found by both `hc` and `gs`. Use `ci.test` to perform a statistical test of (conditional) independence. What do you conclude?

References

[1] Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* **31**(3), 1–22, 2010.