README Final Project: Structure Learning for Gaussian Bayesian Networks

Paul Kahlmeyer

January 24, 2023

1 Problem Setting

In Exercise 3 we had a look at how to learn the parameters of a Bayesian Networks for categoricals. In this setting however we were given a structure in the form of a directed acyclic graph (DAG). Where does this structure come from? This will be the challenge for the final project.

We are looking for a Bayesian Network on Gaussians that has a **good fit** on data and has **few parameters**. So the challenge is to search over the space of DAGs. Unfortunately this space explodes with the number of nodes. More specifically the number of DAGs for n nodes is defined by the super-exponential recurrence

$$G(1) = 1 \tag{1}$$

$$G(n) = \sum_{k=1}^{n} (-1)^{k+1} \binom{n}{k} 2^{k(n-k)} G(n-k) \tag{2}$$

Figure 1 shows the super-exponential growth of the number of DAGs with increasing number of nodes. Hence we cannot fall back on an exhaustive search but have to rely on other search strategies. This is the part where you can get creative.

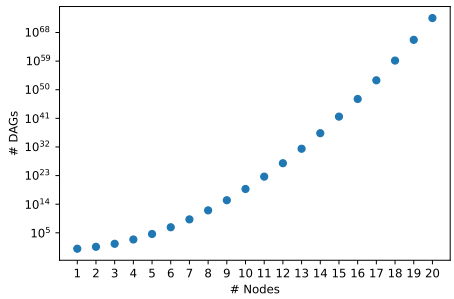

Figure 1: Number of DAGs for increasing number of nodes. This space is super-exponential and it is infeasible to exhaustively search trough it.

2 Task

You will be given a training set of continuous datapoints of which the last dimension is used for classification. Build yourself an environment in which you can assess the Gaussian Bayesian Network for a given adjacency matrix A. Come up with a search strategy and find a sparse Bayesian Network (**few parameters**) that fits well with the data (high likelihood/ **high accuracy** on classification).

You can upload your adjacency matrix to a leaderboard and compare your solution to others. The link will be made available on our Moodle page.

3 About your submission

- At most 10 pages
- Preferrably LaTeX, documentclass article, 10pt.
- Content (exemplary):
  1. Problem Setting
  2. Gaussian Bayesian Networks
  3. Search Strategy
  4. Results

4 Important Dates

- 21.03.23: submission of final report + code via Moodle
- 28.03.23: Oral exam

5 Literature

- Introduction Gaussian Bayesian Networks
- Friedman: Being Bayesian about Network Structure
- Grüttemeier: Efficient Bayesian Network Structure Learning via Parameterized Local Search on Topological Orderings
- Heckerman: A Tutorial on Learning With Bayesian Networks
- Scutari: Who Learns Better Bayesian Network Structures: Accuracy and Speed of Structure Learning Algorithms
- Kitson: A survey of Bayesian Network structure learning

Good luck!