8

Learning

This chapter is devoted to the methods for learning CI structures on the basis of data. However, it is not the aim of this text to provide an overview of numerous methods for learning graphical models from the literature. Rather, the purpose of this chapter is to show how to apply the method of structural imsets to learning DAG models and to indicate that this approach can be extended to learning more general classes of models of CI structure.

Some thoughts about the learning methods which are based on statistical CI tests are mentioned in Section 8.1. The next two sections of the chapter contain an analysis of certain DAG model learning algorithms based on the maximization of a quality criterion, and also deal with related questions. It is argued in Section 8.4 that the method of structural imsets can be applied in this area as well.

8.1 Two approaches to learning

There is plenty of literature about learning graphical models of CI structure – both in the area of statistics and in the area of artificial intelligence. For an overview of some of these methods see § 1.4 and § 4 of Bouckaert [14] and § 3 of Castelo [20]. The literature concerns learning UG models, DAG models and DAG models with hidden variables; most attention is devoted to learning DAG models.

In my view, the algorithms for learning graphical models can be divided into two groups on the basis of the fundamental methodological approach.

- Some of these algorithms are based on *significance tests* between two statistical models (of graphical CI structure). These significance tests often correspond to statistical tests for the validity of some CI statements.
- Other algorithms are based on the maximization of a suitable *quality criterion* designed by a statistician. On the basis of the way to derive the criterion, algorithms of this kind could be further divided into those based

on the Bayesian approach and those which stem from a frequentist point of view.

Nevertheless, some algorithms may be included in both groups because they can be interpreted in both ways and there is a simulation method applicable to learning graphical models which does not belong to either of these two groups (see Remark 8.3 below).

Data faithfulness assumption

Typical examples of algorithms based on significance tests are the SGS algorithm for learning DAG models and its more effective modifications known as the PC algorithm and the PC* algorithm described in Spirtes et al. [122]. These procedures stem from the premise called the *data faithfulness assumption* which can be paraphrased as follows (cf. Section 1.1):

> data are "generated" by a probability measure P which is perfectly Markovian with respect to an object $\boldsymbol{o}$ within the considered universum of objects of discrete mathematics.

In the case of the above algorithms from [122] the universum of the objects is the collection of acyclic directed graphs over N. The algorithms are based on statistical CI tests, that is, tests which – on the basis of given data – either reject or do not reject the hypothesis that a certain elementary CI statement $a \perp\!\!\!\perp b \,|\, C \; [P]$ is true. Tests of this kind are usually derived from statistics which are known to be measures of stochastic conditional dependence (see Section A.9.1 for the concept of a statistic). For example, in § 5.5 of Spirtes et al. [122] two statistics of this kind are mentioned in the discrete case: the X^2-statistic and the G^2-statistic. The goal of the algorithms is to determine the equivalence class of graphs consisting of those acyclic directed graphs with respect to which P is perfectly Markovian. The basic step is an observation that if P is perfectly Markovian with respect to an acyclic directed graph G, then the following characterization of edges and immoralities in G holds:

$$[a,b] \text{ is an edge in } G \quad \Leftrightarrow \quad \forall\, C \subseteq N \setminus \{a,b\} \quad a \not\perp\!\!\!\perp b \,|\, C \; [P]\,,$$

and if $[a,b]$, $[b,c]$ are edges in G while $[a,b]$ is not an edge in G then

$$a \rightarrow c \leftarrow b \text{ in } G \quad \Leftrightarrow \quad \forall\, C \quad c \in C \subseteq N \setminus \{a,b\} \quad a \not\perp\!\!\!\perp b \,|\, C \; [P]$$

– see Verma and Pearl [151] or Kočka et al. [58]. On the condition that data are "generated" from P, the above mentioned statistical CI tests give a criterion for the composite conditional dependence statements on the right-hand side. This allows one to determine the underlying graph of G and all immoralities in G. Thus, a hybrid graph H, called a *pattern*, can be obtained, which has the same underlying graph as G and just those arrows which belong to the immoralities in G; the other edges in H are lines. The final step is a purely