Bayes Networks — Winter 2017/2018
Prof. Dr. R. Kruse / Alexander Dockhorn

**Exercise Sheet 9**

**Exercise 28** Learning from Data

Assume the following conditional independencies between the four attributes $A$, $B$, $C$ and $D$ (as in former exercises, the notation $X \perp\!\!\!\perp Y \mid Z$ states that $X$ is independent of $Y$ given $Z$):

$$A \perp\!\!\!\perp B \mid \emptyset, \qquad A \perp\!\!\!\perp D \mid C, \qquad B \perp\!\!\!\perp D \mid C$$

Assume further that only these independencies as well as those that are deducible by the graphoid axioms (cf. lecture slides) hold true (i.e. the symmetric conuterparts $B \perp\!\!\!\perp A \mid \emptyset$ etc. are satisfied). All other conditional independencies do not hold true. Which conditional independence graph over the four attributes can be read from this information?
(Hint: Remember the special properties of converging edges.)

**Exercise 29** Learning from Data

A simple approach to learn a graphical model from data consists in constructing an optimal spanning tree w.r.t. edge weights that represent the strengths of the attributes connected by that edge. Such a tree is named after its inventors Chow-Liu tree. We consider here the construction of a maximal spanning tree in the relational setting with the Hartley information gain

$$\begin{aligned} I_{\text{gain}}^{(\text{Hartley})}(A,B) &= \log_2\left(\sum_{i=1}^{n_A} R(A=a_i)\right) + \log_2\left(\sum_{j=1}^{n_B} R(B=b_j)\right) \\ &- \log_2\left(\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} R(A=a_i, B=b_j)\right) \\ &= \log_2 \frac{\left(\sum_{i=1}^{n_A} R(A=a_i)\right)\left(\sum_{j=1}^{n_B} R(B=b_j)\right)}{\sum_{i=1}^{n_A}\sum_{j=1}^{n_B} R(A=a_i, B=b_j)}. \end{aligned}$$

as the measure to assess the strength of dependence between attributes $A$ and $B$: Determine for the relation from exercise 13 (repeated below) the Chow-Liu tree w.r.t. the Hartley information gain! Compare the result with the result of exercise 13!

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| $A$ | $a_1$ | $a_1$ | $a_2$ | $a_2$ | $a_2$ | $a_2$ | $a_3$ | $a_3$ |
| $B$ | $b_1$ | $b_1$ | $b_1$ | $b_1$ | $b_3$ | $b_3$ | $b_1$ | $b_2$ |
| $C$ | $c_1$ | $c_2$ | $c_2$ | $c_3$ | $c_2$ | $c_3$ | $c_2$ | $c_2$ |

**Exercise 30** Learning from Data

Consider the following probability distribution:

| | $C = c_1$ | | $C = c_2$ | |
|---|---|---|---|---|
| | $B = b_1$ | $B = b_2$ | $B = b_1$ | $B = b_2$ |
| $A = a_1$ | $4/35$ | $12/35$ | $4/35$ | $1/35$ |
| $A = a_2$ | $1/35$ | $3/35$ | $8/35$ | $2/35$ |

Determine the Chow-Liu tree for that distribution w.r.t. the Shannon information gain

$$I_{\text{gain}}^{(\text{Shannon})}(A, B) = \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} P(A = a_i, B = b_j) \log_2 \frac{P(A = a_i, B = b_j)}{P(A = a_i) \cdot P(B = b_j)},$$

i.e. use the Shannon information gain as the edge weight and determine the maximal spanning tree! Does the result represent a correct decomposition?