Bayesian Networks Winter 2008/2009
Prof. Dr. R. Kruse / M. Steinbrecher

Übungsblatt 7

Aufgabe 22 Lernen aus Daten

Gegeben seien die folgenden bedingten Unabhängigkeiten zwischen den vier Attributen A, B, C und D (wie in früheren Aufgaben bedeute $X \perp\!\!\!\perp Y \mid Z$, daß X unabhängig ist von Y gegeben Z):

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

Nehmen Sie an, daß nur diese und die aus ihnen über die Graphoid-Axiome (siehe Vorlesungsfolien) ableitbaren bedingten Unabhängigkeiten gelten (also z.B. auch die symmetrischen Aussagen $B \perp\!\!\!\perp A \mid \emptyset$ etc.). Alle anderen bedingten Unabhängigkeiten mögen dagegen nicht gelten. Welcher bedingte Unabhängigkeitsgraph über den vier Attributen läßt sich dann aus diesen Informationen ablesen?
(Hinweis: Denken Sie an die besonderen Eigenschaften zusammenlaufender Kanten.)

Aufgabe 23 Lernen aus Daten

Eine sehr einfache Möglichkeit, ein graphisches Modell aus Daten zu lernen, besteht in der Konstruktion eines optimalen spannenden Baumes zu Kantengewichten, die die Stärke der Abhängigkeit der durch die Kante verbundenen Attribute messen. Ein solcher Baum heißt nach den Erfindern dieses Verfahrens auch Chow-Liu-Baum. Wir betrachten in dieser Aufgabe die Konstruktion eines maximalen spannenden Baumes im relationalen Fall, wobei wir den Hartley-Informationsgewinn

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

einsetzen, um die Stärke der Abhängigkeit zwischen zwei Attributen A und B zu messen: Bestimmen Sie für die unten angegebene Relation aus Aufgabe 13 (unten wiederholt) den Chow-Liu-Baum bzgl. des Hartley-Informationsgewinns! Vergleichen Sie das Ergebnis mit der in Aufgabe 13 bestimmten Zerlegung!

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

Aufgabe 24 Lernen aus Daten

Gegeben sei die folgende Wahrscheinlichkeitsverteilung:

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

Bestimmen Sie für diese Verteilung den Chow-Liu-Baum bzgl. des Shannon-Informationsgewinns

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

d.h. benutzen Sie den Shannon-Informationsgewinn als Kantengewicht und bestimmen Sie einen maximalen spannenden Baum bzgl. dieses Kantengewichts! Ist das Ergebnis eine korrekte Zerlegung der Verteilung?