# Maschinelles Lernen II

3. Übung

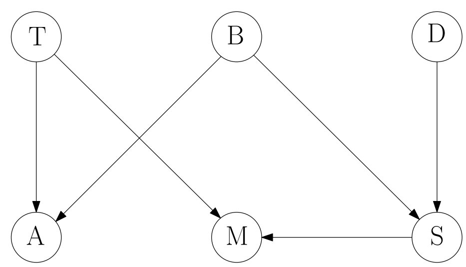
Prof. Tobias Scheffer
Dr. Niels Landwehr
Matthias Bussas

Sommer 2014

Ausgabe am: 29.04.14
Besprechung am: 06.05.14

**Aufgabe 1** *Bedingte Unabhängigkeit und D-Separation*

Wir betrachten die selbe Domäne wie in Aufgabe 1 auf dem letzten Übungsblatt 2. Das heißt wir haben einen Tank (T), einen Starter (S), einen Motor (M), eine Anzeige (A) und eine Batterie (B). Außerdem ist es möglich, dass der Starter defekt (D) ist. Die Faktorisierung der gemeinsamen Verteilung dieser sechs Zufallsvariablen sei durch folgendes graphisches Modell definiert:




Prüfen Sie nun anhand des D-Separation Kriterium, ob die folgenden Unabhängigkeiten in der durch den Graph repräsentierten Verteilung gelten:

- $StarterDefekt \perp Motor \mid Batterie$
- $Batterie \perp Tank \mid Motor$
- $Batterie \perp Motor \mid Starter$
- $Tank \perp StarterDefekt \mid Motor$

Erläutern Sie für jede dieser Unabhängigkeiten auch anschaulich, warum diese gilt bzw. nicht gilt.

**Aufgabe 2** *Trennmengen*

Es sei $G$ ein beliebiger azyklischer Graph und $n$ ein beliebiger Knoten aus $G$. Wir untersuchen nun die Fragestellung, welche Menge von Knoten $M$ wir in dem Graphen $G$ beobachten müssen, damit alle restlichen Knoten von dem Knoten $n$ unabhängig sind. Die kleinste solche Menge nennen wir Trennmenge des Punktes $n$ im Graphen $G$ und bezeichnen sie mit $M(n)$. $M(n)$ ist also die kleinste Menge von Knoten aus $G$ die $n$ nicht enthält und für die gilt:

$$n' \perp n \mid M(n) \qquad \text{für alle } n' \in G \setminus \{n, M(n)\}.$$

Geben Sie eine einfache Beschreibung der Trennmenge eines Knoten $n$ in einem beliebigen Graphen $G$ an und beweisen Sie ihre Richtigkeit.
*Hinweis:* Benutzen Sie das D-Separation Kriterium um diese Menge zu finden und ihre Minimalität zu beweisen.

**Aufgabe 3** *Repräsentation von bedingten Verteilungen*

In der letzten Übung wurde ein modifiziertes Naïve Bayes Modell diskutiert, dass die Verteilung über Merkmale und die Klasse durch

$$p(\mathbf{x}, y) = \prod_{i=1}^{n} p(x_i) p(y|x_1, ..., x_n)$$

modelliert, wobei

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{pmatrix}$$

der Merkmalsvektor ist.

1. Nehmen Sie $n = 3$ an, und die bedingte Verteilung

| $p(y = 1 \| x_1, x_2, x_3)$ | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| 1/2 | 0 | 0 | 0 |
| 2/3 | 1 | 0 | 0 |
| 3/4 | 0 | 1 | 0 |
| 6/7 | 1 | 1 | 0 |
| 4/5 | 0 | 0 | 1 |
| 8/9 | 1 | 0 | 1 |
| 12/13 | 0 | 1 | 1 |
| 24/25 | 1 | 1 | 1 |

Bekanntlich hat dieses Modell das Problem, dass die Verteilung $p(y \mid x_1, ..., x_n)$ exponentiell viele Parameter enthält. Es ist allerdings auch nicht immer nötig, eine solche bedingte Verteilung explizit als Tabelle zu repräsentieren. Die oben dargestellte Verteilung zeigt gewisse Regelmässigkeiten, und kann viel einfacher mit nur drei Parametern charakterisiert werden.

Geben Sie eine Repräsentation der Verteilung an, die mit drei Parametern auskommt. Hinweis: die Sigmoid-Funktion, gegeben durch

$$\sigma(z) = \frac{1}{1 + exp(-z)}$$

könnte hier hilfreich sein, ebenso wie lineare Modelle.

2. Nehmen wir an, wir möchten eine Verteilung $p(y \mid x_1, ..., x_n)$ repräsentieren, die nicht die Regelmässigkeiten der oben gegeben Verteilung zeigt. Wir wollen trotzdem nur wenige Parameter im Modell haben, und sind dafür bereit, die Verteilung nur approximativ zu repräsentieren. Wie könnte man in diesem Fall vorgehen? Hinweis: maschinelles Lernen.