

Bayesian Networks - Bonuszettel

Nicolas Schilling
schilling@ismll.de

December 17, 2013

Abzugeben bis **Dienstag, 7. Januar 18:00**

Exercise 1: Naive Bayes Classifier (10 Punkte)

Bei dem naiven Bayes Klassifikator ist von einer gegebenen Menge von Variablen nur eine von Interesse. Sei I die Variable von Interesse, die anderen Variablen werden als X_1, \dots, X_n bezeichnet, dann sind wir besonders interessiert in:

$$P(I | X_1, \dots, X_n)$$

Nehmen wir an, I modelliert, ob wir vom Weihnachtsmann viele Geschenke bekommen, dann könnten für $n = 5$ die X_1, \dots, X_5 wie folgt definiert sein:

- X_1 : Hilfe im Haushalt $\in \{\text{ja, nein}\}$
- X_2 : Gutes Benehmen $\in \{\text{ja, nein}\}$
- X_3 : Laune des Weihnachtsmanns $\in \{\text{gut, mittel, schlecht}\}$
- X_4 : Müdigkeit der Rentiere $\in \{\text{wach, geht so, müde}\}$
- X_5 : Geldbeutelgröße der Eltern $\in \{\text{groß, mittel, klein}\}$

Als besonderes Merkmal des Bayes Klassifikators gilt die folgende Unabhängigkeitsrelation: Gegeben der Variable von Interesse sind alle X_i voneinander unabhängig:

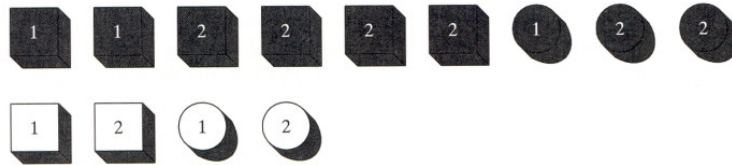
$$P(X_j | I, X_i, X_k) = P(X_j | I, X_i) = P(X_j | I)$$

Das bedeutet, dass jedes Eigenschaftsmerkmal X_i unabhängig von den anderen Variablen X_j die Variable von Interesse I beeinflusst.

- a) Faktorisieren Sie $P(I | X_1, \dots, X_n)$ mit Hilfe der Kettenregel!
- b) Vereinfachen Sie die Formel aus a) indem Sie die Unabhängigkeiten hineinbringen!
- c) Ist das resultierende Modell ein Bayessches Netzwerk? Falls ja, skizzieren Sie den Graphen. Falls nicht, begründen Sie Ihre Antwort!
- d) Was sind die Vorteile des Modells, was sind die Nachteile?

Exercise 2:

Es sei $P(C, L, S)$ die gemeinsame Verteilung der Variablen Farbe (C), Label (L) und Form (S), für das Ziehen eines Objekts mit Zurücklegen. Die Verteilung der Objekte ist auf dem folgenden Bild gegeben:



Berechnen Sie ein Bayessches Netzwerk, welches die gemeinsame Verteilung P faktorisiert, indem Sie

- Die Randverteilungen $P(L), P(C), P(S)$, sowie die gemeinsamen Verteilungen $P(L, S), P(C, S), P(S, L)$ berechnen.
- Alle bedingten Verteilungen jeder Variable berechnen (z.B. für L : $P(L | S), P(L | C)$ und $P(L | C, S)$)
- Das Netzwerk mit gegebenenfalls Unabhängigkeiten zusammenbauen.

Fröhliche Weihnachten!!

