Maschinelles Lernen Blatt 5

Nikolas Zeitler, Joshua Hartmann, Alexander Diegel

June 22, 2016

1 Fragen zur Vorlesung

a) Knoten und Kanten im Bayesschen Netz:
Knoten repräsentieren Wahrscheinlichkeiten von Zufallsvariablen. Kanten repräsentieren kausale Abhängigkeiten (conditional probabilities). Wobei der Knoten von dem der Pfeil weggeht das bereits eingetretene Ereignis ist.

b) Konditionierung: Es interessieren nur einzelne Reihen oder Spalten aus Wahrscheinlichkeitstabellen. Wird bekannt, dass eine Zufallsvariable einen bestimmten Wert annimmt, kann die Verteilung durch Marginalisierung der Tabelle nach der Ebene, bei der die Zufallsvariable den entsprechenden Wert hat, berechnet werden.

c) Inferenz: Ist eine oder mehrere Variablen bekannt, ist man an der Verteilung einer/mehrerer unbekannten variablen interessiert. Das Inferenzproblem ist NP-schwer.

d) Marginalisierung integriert über Variablen, die nicht von Interesse sind. Auf diese weise kann der Zufallsvektor reduziert werden.

2 Bayessche Netze – Berechnung

(a) $P(X = Lachs)$
$= \sum_i \sum_j \sum_k \sum_l (P(a_i) * P(b_j) * P(x_1|a_i, b_j) * P(c_k|x_1) * P(d_l|x_1)$
$= ((0.25 * 0.6) + (0.5 + 0.6 + 0.4 + 0.2) + (0.25 * 0.4) * (0.7 + 0.8 + 0.1 + 0.3)) * 1 * 1$
$= 0.255 + 0.19 = 0.445$

(b) $P(A = Winter, B = Norden, X = Lachs, C = hell, D = duenn)$
$= P(a_1) * P(b_1) * P(x_1|a_1, b_1) * P(c_1|x_1) * P(d_2|x_1)$
$= 0.25 * 0.6 * 0.5 * 0.6 * 0.7 = 0.0315$

(c) $P(X = Lachs|C = Mittel) = P(X = Lachs, C = Mittel)/P(C = Mittel)$
$1)P(X = Lachs, C = Mittel)$
$= \sum_i \sum_j \sum_k (P(a_i) * P(b_j) * P(x_1|a_i, b_j) * P(c_2|x_1) * P(d_k|x_1)$
$= 0.089$

2)$P(C = Mittel)$
$= \sum_i \sum_j \sum_k \sum_l (P(a_i) * P(b_j) * P(x_k|a_i, b_j) * P(c_2|x_k) * P(d_l|x_k)$
$= 0.2555$
$P(X = Lachs|C = Mittel) = 0.089/0.2555 = 0.3483$

(d) $P(A = Sommer|D = breit, X = Barsch) = ...$ wie in c) vorgehen
$= 0.108/0.333 = 0.3243$

(e) $P(C = hell|D = duenn, X = Lachs) = ...$ wie in c) vorgehen
$= 0.1869/0.3115 = 0.6$
–das ist auch plausibel weil D und C unabhängig sind und es für C egal ist, ob D dick oder dünn war.

3 Bayessche Netze – Programmierung

4 Fragen zur Vorlesung

(a) Die Ordnung eines HMM beschreibt die Anzahl an vorherigen Zuständen, von denen der aktuelle Zustand abhängt.

(b) Die Matrix A enthält die Werte für a_{ij}. Diese geben die Übergangswahrscheinlichkeit von ω_i zu ω_j an.

Die Matrix B enthält die Werte für b_{jk}. Diese geben die Wahrscheinlichkeit an, dass Zustand ω_j sich in Visible State v_k äußert.

(c) Die Menge Ω beschreibt die Menge der möglichen Zustände ω_i.

Die Menge V beschreibt die Menge der Visible States v_i.

(d) Um was geht es beim *Bewertungsproblem*?
Hier hat man alle Zustandsübergangs Wahrscheinlichkeiten a_{ij} und b_{ij} sowie alle *visible states* V. Gesucht ist nun die Wahrscheinlichkeit das eine bestimmte Sequenz von *visible states* mit dem Model $\Theta = a_{ij}$ *und* b_{ij} erzeugt wurde. Also $P(V^T|\Theta)$ gesucht

(e) **Forward Algorithm** Berechnet die Wahrscheinlichkeit das man sich durch eine gegebene Sequenz zum Zeitpunkt t in einem versteckten Zustand aufhält. **Backward Algorithm** Berechnet die Wahrscheinlichkeit das man zum Zeitpunkt t eine bestimmte Sequenz beobachtet.

(f) Es macht keinen Unterschied den **Viterbi-Dekoder** mit dem Logarithmus zu berechnen weil hier das Maximum bzw das Optimum einer Funktion gesucht wird. Dieses wird durch den Logarithmus nicht verändert. Dafür fällt das Ableiten leichter.