

WSI für Informatik an der Karls-Eberhardt Universität Tübingen

Machine Learning

Übungsblatt 6

Lea Bey - Benjamin Çoban - Thomas Stüber

29. Juni 2017

6

Aufgabe 6.1.

a) Bayessche Netze

Knoten entsprechen den Zufallsvariablen

Kanten entsprechen den bedingten Wahrscheinlichkeiten zwischen zwei Variablen

b) Die Marginalisierung wird verwendet, um die Wahrscheinlichkeit einer Zufallsvariable durch Aufsummierung über alle bedingten Wahrscheinlichkeiten zu berechnen. Beispiel aus der VL:

$$p(A) = \sum_i \left(\sum_j p(A, b_i, c_j) \right)$$

Lerne ich eine neue Zufallsvariable kennen, kann ich das weiterverarbeiten: Das Conditioning berechnet nun zu gegebener $p(A, B, C)$ die bedingte Wahrscheinlichkeit

$$p(B, C \mid a) = \frac{p(B, C, a)}{p(a)}$$

wenn $p(a)$ bekannt. Die Inferenz steht zur Verfügung, um etwas über die Elternknoten in Erfahrung zu bringen, sobald wir die Kinderknoten auswerten können. Die Begriffe gehen thematisch Hand in Hand. Um zB die Inferenz zu berechnen, benötigen wir oft die Marginalisierung.

c) HMM

Markov Modelle haben zu einem Zeitpunkt t einen Modellzustand ω_t . Die Ordnung beschreibt, von welchen vorherigen Zuständen der Zustand ω_{t+1} abhängt. Ein Markov Modell erster Ordnung besagt also, dass ω_{t+1} alleine von ω_t abhängt.

d) zusammen mit

e) Das HMM entspricht einem Automaten, welcher eine Menge Ω an Zuständen hat. Diese sind nicht einsichtbar. Die Übergänge a_{ij} sind mit der Wahrscheinlichkeit gewichtet, wie wahrscheinlich es ist, dass von Zustand ω_j zu ω_i gewechselt wird. Das entspricht den Einträgen der Matrix A . Um auch etwas auswerten zu können, gibt es eine Ausgabemenge V , welcher zum Zustandwechsel mit einer gewissen Wahrscheinlichkeit b_{jk} ausgegeben wird. Das entspricht den Einträgen der Matrix B . Haben wir also eine Sequenz V^T , können wir diese dann mit verschiedenen Algorithmen auswerten, um dann nachzuvollziehen, welche Zustandswechsel meines HMM zur Sequenz am wahrscheinlichsten ist.

f) Betrachten wir die drei grundlegenden Probleme zum HMM:

- Evaluationsproblem

Ein HMM mit den Matrizen A, B ist vollständig vorhanden. Nun geht es um die Berechnung, wie wahrscheinlich es zu einer Sequenz V^T ist, dass dieses Modell diese Sequenz erzeugt hat.

- Dekodierungsproblem

Ein HMM ist mit einer Sequenz V^T vorhanden. Was ist die wahrscheinlichste Sequenz an (versteckten) Zuständen ω^T , sodass V^T daraus abgeleitet wurde?

- Lernproblem

Wir haben ein HMM, aber nur die jeweilige Anzahl der versteckten und Auswertungszustände $(|\Omega|, |V|)$ und wollen nun die Parametermatrizen A, B aus einem Satz an Trainingssequenzen berechnen.

- g) Mit dem Forward- oder Backward-Algorithmus lässt sich sowohl das Evaluationsproblem, als auch das Dekodierungsproblem algorithmisch mit vernünftiger Laufzeit angehen. Naiv müssten wir jede mögliche Zustandsfolge evaluieren, was in exponentieller Laufzeit liegt. Mit den Forward- und Backward Algorithmus arbeiten wir allerdings rekursiv auf das höchstwahrscheinlichste nächste Element der Folge an Zuständen (Dekodierungsproblem), was deutlich an Rechenleistung spart.
- h) Der Logarithmus ist streng monoton steigend und minimiert entstehende numerische Fehler. Des weiteren erhält er die Rechenoperationen, da in der einzelnen Summe exponiert wird.

Aufgabe 6.2.

Betrachte das Bayessche Netz mit den verteilten Wahrscheinlichkeiten auf dem Übungsblatt. Zu berechnen sind folgende Wahrscheinlichkeiten:

- a) $p(x_1)$

$$p(x_1) = \sum_{i,j,k,l} p(x_1, a_i, b_j, c_k, d_l) \quad (6.1)$$

$$\underbrace{=}_{\text{Graph } i,j,k,l} \sum p(x_1 \mid a_i, b_j) \cdot p(a_i) \cdot p(b_j) \cdot p(c_k \mid x_1) \cdot p(d_l \mid x_1) \quad (6.2)$$

$$= \sum_{i,j} \sum_{k,l} p(x_1 \mid a_i, b_j) \cdot p(a_i) \cdot p(b_j) \cdot p(c_k \mid x_1) \cdot p(d_l \mid x_1) \quad (6.3)$$

In Zeile 6.3 betrachten wir nun $p(c_k \mid x_1)$ und $p(d_l \mid x_1)$. Dies heißt, dass wir über eine normierte Zeile iterieren, sowohl $p(c_k \mid x_1)$ als auch $p(d_l \mid x_1)$ summieren sich zu 1 auf, wenn wir sie ausklammern. Somit fällt die innere Summe mit den einzelnen Wahrscheinlichkeiten raus. Übrig bleibt also:

$$= \sum_{i,j} p(x_1 \mid a_i, b_j) \cdot p(a_i) \cdot p(b_j) \quad (6.4)$$

Macht auch Sinn, denn es interessiert an Hand vom Graphen für die Wahrscheinlichkeit von Lachs nicht, ob er o.B.d.A. hell oder breit ist.

$$= p(a_i) \cdot \sum_{i,j} p(x_1 \mid a_i, b_j) \cdot p(b_j) \quad (6.5)$$

$$= 0,25 \cdot (0,5 \cdot 0,6 + 0,7 \cdot 0,4 + 0,6 \cdot 0,6 + 0,8 \cdot 0,4) \quad (6.6)$$

$$+ 0,4 \cdot 0,6 + 0,1 \cdot 0,4 + 0,2 \cdot 0,6 + 0,3 \cdot 0,4) = \underline{0,445} \quad (6.7)$$

b) $p(a_1, b_1, x_1, c_1, d_2)$

$$= p(a_1) \cdot p(b_1) \cdot p(x_1 | a_1, b_1) \cdot p(c_1 | x_1) \cdot p(d_2 | x_1) \\ = 0,25 \cdot 0,6 \cdot 0,5 \cdot 0,6 \cdot 0,7 = 0,0315$$

c) $p(x_1 | c_2)$

Berechne zuerst $p(c_2)$:

$$p(c_2) = \sum_{i,j,k,l} p(a_i, b_j, x_k, c_2, d_e)$$

Da wir in Teilaufgabe 1 bereits die Wahrscheinlichkeit für Lachs - und somit auch für Barsch - berechnet haben, kann ein erheblicher Teil der Rechnung übersprungen werden. Das liegt am Bayesschen Netz: Die Wahrscheinlichkeit eines Knotens ist bedingt durch seine Elternknoten - in unserem Fall X und bereits ausgerechnet.

$$= \sum_i p(c_2 | x_i) \cdot p(x_i) \\ = 0,2 \cdot 0,445 + 0,3 \cdot 0,555 = 0,2555$$

$$p(x_1 | c_2) = \frac{p(c_2, x_1)}{p(c_2)} \underbrace{=}_{\text{Satz von Bayes}} \frac{p(x_1) \cdot p(c_2 | x_1)}{p(c_2)} = 0,3483$$

d) $p(a_3 | x_2, d_1)$ Laut Tutorium gilt folgende Gleichung:

$$p(a_3 | x_2, d_1) = \frac{p(a_3, x_2, d_1)}{p(x_2) \cdot p(d_1 | x_2)} \quad (6.8)$$

Berechne nun den Zähler aus 6.8

$$p(a_3, x_2, d_1) = \sum_{i,j} p(a_3) \cdot p(b_i) \cdot p(x_2 | a_3, b_i) \cdot \underbrace{p(c_j | x_2)}_{=1} \\ = p(a_3) \cdot p(d_1 | x_2) \cdot \sum_i p(b_i) \cdot p(x_2 | a_3, b_i) \\ = 0,25 \cdot 0,6 \cdot (0,6 \cdot 0,6 + 0,4 \cdot 0,9) = 0,108$$

Setzen wir das ganze nun in 6.8 ein, erhalten wir:

$$p(a_3 | x_2, d_1) = \frac{0,108}{0,555 \cdot 0,6} = 0,3243$$

e) $p(c_1 | d_2, x_1)$

Das Vorgehen ist gleich wie bei Teilaufgabe d)

$$p(x_1, c_1, d_2) = \sum_{i,j} p(a_i) \cdot p(b_j) \cdot p(x_1 | a_i, b_j) \cdot p(c_1 | x_1) \cdot p(d_2 | x_1) \\ = p(c_1 | x_1) \cdot p(d_2 | x_1) \cdot \underbrace{\sum_{i,j} p(a_i) \cdot p(b_j) \cdot p(x_1 | a_i, b_j)}_{=p(x_1)}$$

Einsetzen in 6.8

$$\begin{aligned} p(c_1 \mid d_2, x_1) &= \frac{p(c_1 \mid x_1) \cdot p(d_2 \mid x_1) \cdot p(x_1)}{p(d_2 \mid x_1) \cdot p(x_1)} \\ &= p(c_1 \mid x_1) = 0,6 \end{aligned}$$