Grundlagen des maschinellen Lernens

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Tobias Lang
Mathias Schickel

Andreas Schilling
Sommersemester 2016

Übungsblatt 7

Ausgabe: 23.06.2016; Abgabe: bis 07.07.2016, 23:59 Uhr.

Aufgabe 1 (Hidden Markov Model) (20 Punkte)

Gegeben sei das folgende Hidden Markov Model:

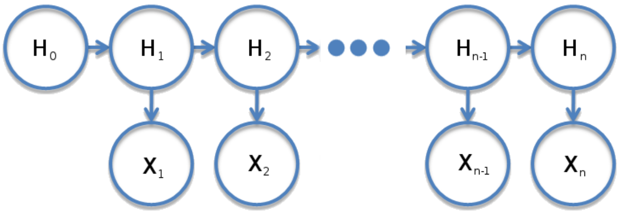

Abbildung 1: Das zu untersuchende Hidden Markov Model

Die Übergangswahrscheinlichkeiten der verborgenen Zustände seien gegeben durch

$$
\begin{array}{c} \\ H_0 = 1 \end{array}
\begin{array}{cc} H_1 = 1 & H_1 = 2 \\ \begin{pmatrix} 0.4 & 0.6 \end{pmatrix} \end{array},
\quad
\begin{array}{c} \\ H_1 = 1 \\ H_1 = 2 \end{array}
\begin{array}{c} H_2 = 1 \quad H_2 = 2 \\ \begin{pmatrix} 0.3 & 0.7 \\ 0.7 & 0.3 \end{pmatrix} \end{array},
\quad
\begin{array}{c} \\ H_2 = 1 \\ H_2 = 2 \end{array}
\begin{array}{c} H_3 = 1 \quad H_3 = 2 \\ \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \end{array},
$$

$$
\begin{array}{c} \\ H_3 = 1 \\ H_3 = 2 \end{array}
\begin{array}{c} H_4 = 1 \quad H_4 = 2 \\ \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix} \end{array},
\quad
\begin{array}{c} \\ H_4 = 1 \\ H_4 = 2 \end{array}
\begin{array}{c} H_5 = 1 \quad H_5 = 2 \\ \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \end{array}
$$

und die der sichtbaren Zustände durch

$$
\begin{array}{c} \\ H_1 = 1 \\ H_1 = 2 \end{array}
\begin{array}{c} X_1 = A \quad X_1 = B \\ \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix} \end{array},
\quad
\begin{array}{c} \\ H_2 = 1 \\ H_2 = 2 \end{array}
\begin{array}{c} X_2 = C \quad X_2 = D \\ \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix} \end{array},
\quad
\begin{array}{c} \\ H_3 = 1 \\ H_3 = 2 \end{array}
\begin{array}{c} X_3 = E \quad X_3 = F \\ \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix} \end{array},
$$

$$
\begin{array}{c} \\ H_4 = 1 \\ H_4 = 2 \end{array}
\begin{array}{c} X_4 = G \quad X_4 = H \\ \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix} \end{array},
\quad
\begin{array}{c} \\ H_5 = 1 \\ H_5 = 2 \end{array}
\begin{array}{c} X_5 = I \quad X_5 = J \\ \begin{pmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{pmatrix} \end{array}.
$$

a) Man erstelle in Matlab eine geeignete Repräsentation des Hidden Markov Models. (Siehe dazu auch die unten stehenden Hinweise.)

b) Man implementiere eine Funktion `forward(X, t, model)`, die zu einer gegebenen Sequenz $X_1, \dots, X_n$ an der t-ten Stelle den *Forward-Algorithmus* auswertet. (Wenn das Modell aus mehreren Variablen besteht, verändert sich die Signatur der Funktion entsprechend.) (4 Punkte)

c) Man implementiere eine Funktion `backward(X, t, model)`, die zu einer gegebenen Sequenz $X_1, \dots, X_n$ an der t-ten Stelle den *Backward-Algorithmus* auswertet. (Für Modelle aus mehreren Variablen gilt derselbe Hinweis wie in der letzten Teilaufgabe.) (4 Punkte)

d) Man implementiere eine Funktion `viterbi(X, model)`, die zu einer gegebenen Sequenz $X_1, \dots, X_n$ den *Viterbi-Algorithmus* auswertet. (Die Ausgabe ist hier keine keine Zahl, sondern ein Pfad.) (8 Punkte)

e) Man bestimme $\mathbb{P}(H_5 = 1 | X = \text{ACFGI})$. (1 Punkt)

f) Man berechne $\mathbb{P}(H_5 = 1 | X = \text{BCEGJ})$. (1 Punkt)

g) Man bestimme die wahrscheinlichste Sequenz $H_1, \dots, H_n$, die die Sequenz $X = \text{ACFGI}$ hervorgerufen haben könnte. (2 Punkte)

Hinweise:

- Ist es für die Rekursion nötig, dann darf den Anfangs- und Endzuständen die Wahrscheinlichkeit 1 zuwiesen werden (wie in der Vorlesung erwähnt).
- Aufgabenteil d) ist nicht trivial, wenn Ihr nicht viel Erfahrung bei der Matlab-Programmierung habt. Überlegt Euch dazu, wie man sich während der Suche nach dem Maximum die Knoten merken kann.
- Um die Abtipparbeit der obigen Übergangswahrscheinlichkeiten zu reduzieren, wird das Matlab-Skript `createHmm.m` zur Verfügung gestellt.

Aufgabe 2 (Fragen zur Vorlesung) (9 Punkte)

Man beantworte die folgenden Fragen:

a) Aus der Vorlesung ist bekannt, dass die Wahrscheinlichkeit eines Punktes x durch den Ausdruck

$$p(x) \approx \frac{k/n}{V}$$

angenähert werden kann. Warum gilt für $n \to \infty$ mit $k_n \to \infty$ und $V_n \to 0$ für den Schätzer $p_n(x)$ die Identität

$$p_n(x) = \frac{k_n/n}{V_n} \to p(x)?$$

b) Was ist die Strategie der Dichteschätzung bei dem *Parzenfenster-Verfahren* und bei dem *Nächster-Nachbar-Verfahren*?

c) Inwiefern handelt es sich bei beiden in der letzten Aufgabenstellung genannten Verfahren um sogenannte *nicht-parametrische* Methoden? Sie werden gelegentlich auch *Prototypen-basierte Verfahren* genannt. Erkläre, weswegen dies sinnvoll ist.

Aufgabe 3 (Parzenfenster) (10 Punkte)

Die Schätzung $p_n(x)$ einer Dichte für eine Normalverteilung bei einem Datensatz mit n Stichproben hat die Gestalt

$$p_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - x)^2}{2\sigma^2}\right).$$

a) Plottet die Punkte aus $D = \{2, 4, 5, 8, 15\}$ und deren Beitrag zur Dichteschätzung. In dieselbe Grafik soll die über den Datensatz D geschätzte Dichtefunktion im Intervall $[0, 20]$ geplottet werden. Für die Fensterfunktion soll dabei $\sigma = 1$ verwendet werden.

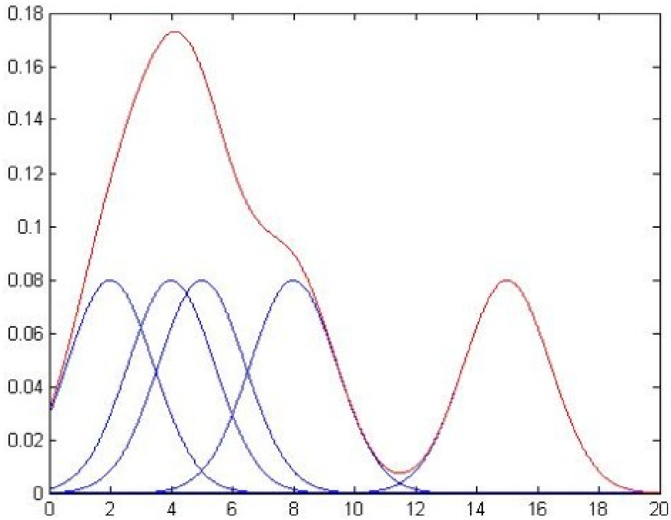

Abbildung 2: Dichteschätzung mit Parzenfenster und dem Beitrag der einzelnen Datenpunkte

b) Man implementiere ein probabilistisches neuronales Netz (PNN) für die Datensätze D_1 und D_2, die als `.csv`-Files vorliegen.

c) Man klassifiziere die Punkte aus der Menge

$$M := \{(0,0), (7,7), (30,20), (14,40)\}$$

nach D_1 und D_2.