Übung zur Vorlesung

# Maschinelles Lernen 2

Sommersemester 2013

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und
Theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

## Exercise Sheet 9

Due **June 17**, 9am local time on ISIS

1. **The Viterbi algorithm (70 + 30 + 15* + 15* P)**

(a) *Implement* the Viterbi algorithm which computes the most likely hidden state sequence for a sequence of observations. For this, write a MATLAB function `viterbi` with signature

```
x = viterbi(A, B, pi, y),
```

which takes the transistion matrix `A`, emission matrix `B`, initial distribution `pi`, and the observed sequence `y`. It outputs the MLE for the hidden state sequence `x`.

(Pseudocode of the algorithm can be found on page 2)

(b) Perform the following experiment: for the Hidden Markov Model of sheet 9, that is,

$$A = \begin{pmatrix} 0{,}1 & 0{,}9 \\ 0{,}5 & 0{,}5 \end{pmatrix} \quad B = \begin{pmatrix} 0{,}2 & 0{,}8 \\ 0{,}4 & 0{,}6 \end{pmatrix} \quad \pi = \begin{pmatrix} 1 & 0 \end{pmatrix},$$

generate sequences of length $5, 10$ and $20$; for each length generate a number of $N = 1000$ pairs of output sequences and hidden state sequences. On these sequences, for each length, compare (i) the Viterbi algorithm and (ii) the algorithm which randomly uniformly estimates a state sequence by (i) plotting for each length, and $n$ between 1 and the length the relative frequency of the algorithm correctly estimating the hidden state at position $n$ (this is three plots, in each two curves), and (ii) for each length, computing the relative frequency of both algorithms succeeding in completely identifying the state sequence correctly (this is three times two numbers).

Hint: if you want to do exercise (d*), and also generally, it is advised that you save the generated sequences for future use. For this, use the MATLAB command `save`.

(c*) In a Markov Model, there are often state sequences resp. distributions to which the model converges. *Prove* for the model from sheet 9 given in exercise (b) that no matter what the initial distribution $\pi$ is, the limit $\pi_\infty := \lim_{n\to\infty} \pi \cdot A^n$ is the same.

(d*) In your comparison in exercise (b), add the algorithm which does not estimate the outputs uniformly, but with independent probabilities $\pi_\infty \cdot B$.

**Viterbi-algorithmus** (cf. Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceedings of the IEEE, Vol. 77, No. 2, 1989, p. 264ff.)

| | | |
|---|---|---|
| Variables: | $n$ | number of hidden states |
| | $m$ | number of possible emissions |
| | $N$ | Length of sequence |
| | $a_{ij}$ | transition probabilities |
| | $b_{ij}$ | probability $P()$ of emitting symbol $j$ in state $i$ |
| | $\pi_i$ | probability to start in state $i$ |
| | $y(t)$ | observation at time $t$ |
| | $\widehat{x}(t)$ | estimated state at time $t$ |
| | $P^*$ | probability for $y(t)$ given $\widehat{x}(t)$ |
| | $\delta_i(t)$ | "best" probability of state $i$ at time $t$ |
| | $\psi_i(t)$ | "best" state $i$ at time $t$ |

Initalization:

$$\begin{aligned} \delta_i(1) &= \pi_i \cdot b_{ix(1)}, \qquad 1 \le i \le n, \\ \psi_i(1) &= 0. \end{aligned}$$

Rekursion:

$$\begin{aligned} \delta_j(t) &= \max_{1 \le i \le n} [\delta_i(t-1) \cdot a_{ij}] \cdot b_{jx(t)}, \qquad 2 \le t \le N,\ 1 \le j \le n \\ \psi_j(t) &= \operatorname*{argmax}_{1 \le i \le n} [\delta_i(t-1) \cdot a_{ij}], \qquad 2 \le t \le N,\ 1 \le j \le n. \end{aligned}$$

Termation:

$$\begin{aligned} P^* &= \max_{1 \le i \le n} [\delta_i(N)], \\ \widehat{x}(T) &= \operatorname*{argmax}_{1 \le i \le n} [\delta_i(N)]. \end{aligned}$$

Backtracking:

$$\widehat{x}(t) = \psi_{\widehat{x}(t+1)}(t+1), \qquad t = N-1, N-2, \cdots, 1.$$

Please ask questions in the ISIS discussion forums for Machine Learning 2:
`https://www.isis.tu-berlin.de/course/view.php?id=8005`