Machine Learning Worksheet 8

HMMs

1 Recognizing words

A word recognizer uses HMMs to represent each word. The speech signal is described at each frame interval by a sequence of acoustic features.

Problem 1. Consider the following particular discrete HMM $\lambda = \{q, A, B\}$ (here, the acoustic features are four discrete values ranging from 1 through 4).:

$$q = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}; \quad A = \begin{pmatrix} 0.3 & 0.7 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{pmatrix}; \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.2 & 0.6 \end{pmatrix}$$

An observed sequence was $\mathbf{x} = (1, 3, 2, 4, 1)$.

(a) Draw the state graph of the Markov chain. Is this a reasonable model for recognizing a word? Why/why not?

(b) Determine the total number of possible state sequences of length $T = 5$ for this particular HMM.

(c) Use the forward algorithm to calculate the probability that the given word model produced the given observation sequence. Show intermediate computation results.

(d) Determine the most probable state sequence for the given observed sequence using the Viterbi algorithm. Visualize your computation steps using a trellis.

2 Non-standard emission models

Problem 2. ⋆ Consider a hidden Markov model in which the emission densities are represented by a parametric model $p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{w})$, such as a linear regression model or a neural network, in which $\boldsymbol{w}$ is a vector of adaptive parameters. *Describe* how the parameters $\boldsymbol{w}$ can be learned from data using maximum likelihood.