CSE 150. Assignment 5 *Winter 2015*

Out: *Tue Feb 18*
Due: *Tue Feb 24*

Supplementary reading:

- Russell & Norvig, Chapter 15.
- L. R. Rabiner (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77(2):257–286.

5.1 Viterbi algorithm

In this problem, you will decode an English phrase from a long sequence of non-text observations. To do so, you will implement the same algorithm used in modern engines for automatic speech recognition. In a speech recognizer, these observations would be derived from real-valued measurements of acoustic waveforms. Here, for simplicity, the observations only take on binary values, but the high-level concepts are the same.

Consider a discrete HMM with $n = 26$ hidden states $S_t \in \{1, 2, \ldots, z\}$ and binary observations $O_t \in \{0, 1\}$. Download the ASCII data files from the course web site for this assignment. These files contain parameter values for the initial state distribution $\pi_i = P(S_1 = i)$, the transition matrix $a_{ij} = P(S_{t+1} = j | S_t = i)$, and the emission matrix $b_{ik} = P(O_t = k | S_t = i)$, as well as a long bit sequence of $T = 45000$ observations.

Use the Viterbi algorithm to compute the most probable sequence of hidden states conditioned on this particular sequence of observations. As always, you may program in the language of your choice. Turn in the following:

(a) a hard-copy print-out of your source code

(b) a plot of the most likely sequence of hidden states versus time.

To check your answer: suppose that the hidden states $\{1, 2, \ldots, 26\}$ represent the letters $\{a, b, \ldots, z\}$ of the English alphabet. If you have implemented the Viterbi algorithm correctly, the most probable sequence of hidden states (ignoring repeated elements) will reveal a highly recognizable phrase.

5.2 Conditional independence

Indicate the **smallest** subset of evidence nodes that must be considered to compute each conditional probability shown below. The first two problems are done as examples. (You may assume everywhere that $2 < t < T-1$: i.e., do not worry about special boundary cases.)

$$P(S_t|S_1, S_2, \ldots, S_{t-1}) = P(S_t|S_{t-1})$$

$$P(O_t|S_1, S_2, \ldots, S_T) = P(O_t|S_t)$$

$$P(S_t|S_{t+1}, S_{t+2}, \ldots, S_T) = \underline{\hspace{8cm}}$$

$$P(S_t|O_t, O_{t-1}, O_{t+1}) = \underline{\hspace{8cm}}$$

$$P(S_t|O_t, O_{t-1}, O_{t+1}, S_{t-1}, S_{t+1}) = \underline{\hspace{8cm}}$$

$$P(S_t|S_1, S_T, O_1, O_t, O_T) = \underline{\hspace{8cm}}$$

$$P(S_t|O_t, O_{t+1}, \ldots, O_T) = \underline{\hspace{8cm}}$$

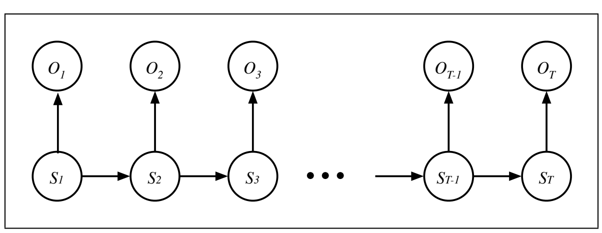
$$P(O_t|O_1, O_2, \ldots, O_{t-1}) = \underline{\hspace{8cm}}$$

$$P(O_t|O_1, O_2, \ldots, O_{t-1}, S_{t-1}) = \underline{\hspace{8cm}}$$

$$P(O_t|O_1, O_2, \ldots, O_{t-1}, S_{t-2}) = \underline{\hspace{8cm}}$$

$$P(O_t|S_{t-2}, S_{t-1}, S_{t+1}, S_{t+2}) = \underline{\hspace{8cm}}$$

$$P(O_t|O_{t-1}, O_{t+1}, S_1, S_T) = \underline{\hspace{8cm}}$$



5.3 Inference in HMMs

Consider a discrete HMM with hidden states S_t, observations O_t, transition matrix $a_{ij} = P(S_{t+1} = j | S_t = i)$ and emission matrix $b_{ik} = P(O_t = k | S_t = i)$. In class, we defined the forward-backward probabilities:

$$\begin{aligned} \alpha_{it} &= P(o_1, o_2, \ldots, o_t, S_t = i), \\ \beta_{it} &= P(o_{t+1}, o_{t+2}, \ldots, o_T | S_t = i), \end{aligned}$$

for a particular observation sequence $\{o_1, o_2, \ldots, o_T\}$ of length T. This problem will increase your familiarity with these definitions. In terms of these probabilities, which you may assume to be given, as well as the transition and emission matrices of the HMM, show how to (efficiently) compute the following probabilities:

(a) $P(S_t = i | S_{t+1} = j, o_1, o_2, \ldots, o_T)$

(b) $P(S_{t+1} = j | S_t = i, o_1, o_2, \ldots, o_T)$

(c) $P(S_{t+1} = k | S_{t-1} = i, S_t = j, o_1, o_2, \ldots, o_T)$

(d) $P(S_{t+1} = k | S_{t-1} = i, o_1, o_2, \ldots, o_T)$

In all these problems, you may assume that $t > 1$ and $t < T$; in particular, you are *not* asked to consider the boundary cases.