CSE 150. Assignment 5

Out: *Tue Feb 19*
Due: *Tue Feb 26*

Recommended 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 $T = 60000$ observations.

Use the Viterbi algorithm to compute the most probable sequence of hidden states conditioned on this particular sequence of observations. **Turn in a print-out of your source code, as well as a plot of the most likely sequence of hidden states versus time.** You may program in the language of your choice.

To check your answer: suppose that the hidden states $\{1, 2, \ldots, 26\}$ represent the letters $\{a, b, \ldots, z\}$ of the English alphabet. The most probable sequence of hidden states (ignoring repeated elements) will reveal a highly recognizable phrase of 2012.

5.2 Conditional independence

Consider the hidden Markov model (HMM) shown below, with hidden states S_t and observations O_t for times $t \in \{1, 2, \ldots, T\}$. State whether the following statements of conditional independence are true or false.

__________ $P(S_{t+1}|S_t) = P(S_{t+1}|S_t, O_t)$

__________ $P(S_t|S_{t-1}) = P(S_t|S_{t-1}, S_{t+1})$

__________ $P(S_{t+1}|S_t) = P(S_{t+1}|S_t, O_{t+1})$

__________ $P(S_t|O_t) = P(S_t|O_1, O_2, \ldots, O_t)$

__________ $P(O_{t+1}|S_t) = P(O_{t+1}|S_t, O_t)$

__________ $P(O_t|O_{t-1}) = P(O_t|O_1, O_2, \ldots, O_{t-1})$

__________ $P(O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(O_t|O_1, \ldots, O_{t-1})$

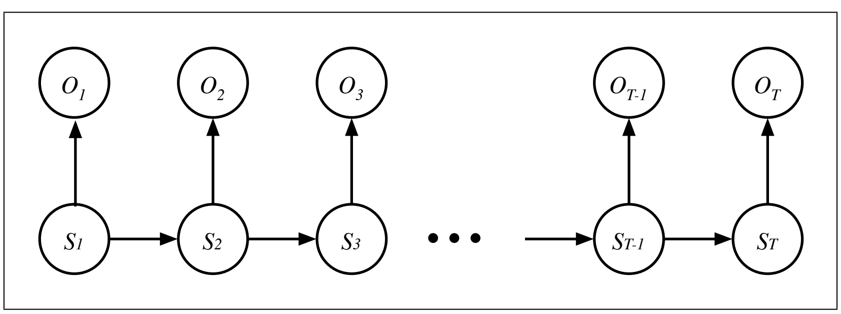
__________ $P(S_2, S_3, \ldots, S_T|S_1) = \prod_{t=2}^{T} P(S_t|S_{t-1})$

__________ $P(S_1, S_2, \ldots, S_{T-1}|S_T) = \prod_{t=1}^{T-1} P(S_t|S_{t+1})$

__________ $P(O_1, O_2, \ldots, O_T|S_1, S_2, \ldots, S_T) = \prod_{t=1}^{T} P(O_t|S_t)$

__________ $P(S_1, S_2, \ldots, S_T, O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t, O_t)$

__________ $P(S_1, S_2, \ldots, S_T|O_1, O_2, \ldots, O_T) = \prod_{t=1}^{T} P(S_t|O_t)$



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. 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 posterior probability:

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

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