## CSE 250A. Assignment 7

**Out:** *Tue Nov 17*
**Due:** *Tue Nov 24* (2 pm, Pacific Time, via Gradescope)
**Reading:** Russell & Norvig, Chapter 15.

### 7.1 Viterbi algorithm

In this problem, you will decode an English sentence from a long sequence of non-text observations. To do so, you will implement the same basic algorithm used in most 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 $n = 27$ hidden states $S_t \in \{1, 2, \ldots, 27\}$ 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 $T = 175000$ 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 $\{\mathtt{a}, \mathtt{b}, \ldots, \mathtt{z}\}$ of the English alphabet, and suppose that hidden state 27 encodes a space between words. If you have implemented the Viterbi algorithm correctly, the most probable sequence of hidden states *(ignoring repeated elements)* will reveal a recognizable message.

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

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

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

### 7.3 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.

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

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

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

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

__________ $P(O_t|S_{t-1}) = P(O_t|S_{t-1}, O_{t-1})$

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

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

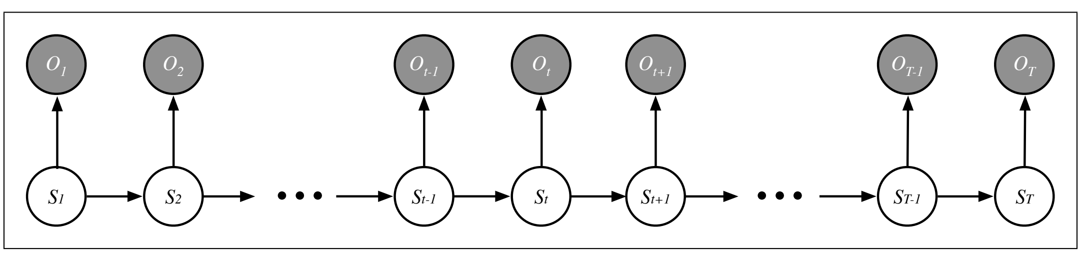
__________ $P(S_2, S_3, \ldots, S_T|S_1) = \prod_{t=2}^{T} P(S_t|S_{t-1})$

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

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

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

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



## 7.4 Belief updating

In this problem, you will derive recursion relations for real-time updating of beliefs based on incoming evidence. These relations are useful for situated agents that must monitor their environments in real-time.

(a) Consider the discrete hidden Markov model (HMM) with hidden states $S_t$, observations $O_t$, transition matrix $a_{ij}$ and emission matrix $b_{ik}$. Let

$$q_{it} \;=\; P(S_t\!=\!i|o_1, o_2, \ldots, o_t)$$

denote the conditional probability that $S_t$ is in the $i^{\text{th}}$ state of the HMM based on the evidence up to and including time $t$. Derive the recursion relation:

$$q_{jt} = \frac{1}{Z_t} b_j(o_t) \sum_i a_{ij} q_{it-1} \quad \text{where} \quad Z_t = \sum_{ij} b_j(o_t) a_{ij} q_{it-1}.$$

Justify each step in your derivation—for example, by appealing to Bayes rule or properties of conditional independence.

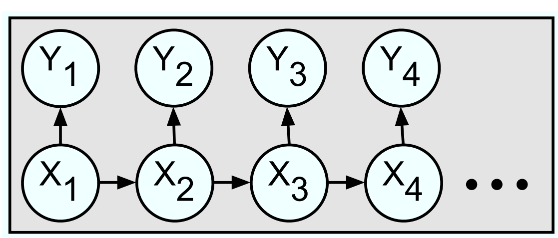
(b) Consider the dynamical system with *continuous, real-valued* hidden states $X_t$ and observations $Y_t$, represented by the belief network shown below. By analogy to the previous problem (replacing sums by integrals), derive the recursion relation:

$$P(x_t|y_1, y_2, \ldots, y_t) \;=\; \frac{1}{Z_t} P(y_t|x_t) \int dx_{t-1}\, P(x_t|x_{t-1})\, P(x_{t-1}|y_1, y_2, \ldots, y_{t-1}),$$

where $Z_t$ is the appropriate normalization factor,

$$Z_t = \int dx_t\, P(y_t|x_t) \int dx_{t-1}\, P(x_t|x_{t-1})\, P(x_{t-1}|y_1, y_2, \ldots, y_{t-1}).$$

In principle, an agent could use this recursion for real-time updating of beliefs in arbitrarily complicated continuous worlds. In practice, why is this difficult for all but Gaussian random variables?



## 7.5 V-chain

Consider the belief network shown below with hidden nodes $X_t$ and $Y_t$ and observed nodes $O_t$. Suppose also that the same CPTs are shared across nodes at different times, with parameters

$$
\begin{aligned}
\pi_j &= P(Y_t = j), \\
a_{j\ell} &= P(X_{t+1} = \ell | Y_t = j), \\
b_{ij}(k) &= P(O_t = k | X_t = i, Y_t = j),
\end{aligned}
$$

that are valid for all times $t$. In this problem you will derive an efficient algorithm for computing the likelihood $P(o_1, o_2, \ldots, o_T)$ in terms of these parameters.

(a) **Base case**

Show how to compute the *marginal* probability $P(Y_1 = j, O_1 = o_1)$ in terms of the prior probability $P(X_1 = i)$ and the parameters $\{\pi, b\}$ of this model.

(b) **Forward algorithm**

Consider the probabilities $\alpha_{jt} = P(o_1, o_2, \ldots, o_t, Y_t = j)$ for a particular sequence of observations $\{o_1, o_2, \ldots, o_T\}$. Derive an efficient recursion to compute these probabilities at time $t+1$ from those at time $t$ and the parameters $\{\pi, a, b\}$ of this model. Justify each step in your derivation.

(c) **Likelihood**

Show how to compute the likelihood $P(o_1, o_2, \ldots, o_T)$ efficiently in terms of the marginal probabilities $\alpha_{jT}$ as defined in part (b).

(d) **Complexity**

Suppose that $X_t \in \{1, 2, \ldots, n_x\}$ and $Y_t \in \{1, 2, \ldots, n_y\}$. How does the computation of the likelihood scale in the length of the sequence $T$ and the cardinalities $n_x$ and $n_y$ of the hidden states? Justify your answer.