ECE 417 Multimedia Signal Processing
Solutions to Homework 4

UNIVERSITY OF ILLINOIS
Department of Electrical and Computer Engineering

Assigned: Tuesday, 10/10/2023; Due: Tuesday, 10/17/2023
Reading: L.R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, 1989

Problem 4.1

In a first-order Markov model, the state at time t depends only on the state at time $t-1$. A **second-order Markov model** is a model in which the state at time t depends on a short list of recent states. For example, consider a model in which q_t depends on the most recent **two** frames. Let's suppose the model is fully defined by the following three types of parameters:

- **Initial segment probability:** $\pi_{i,j} \equiv \Pr\{q_1 = i, q_2 = j|\Lambda\}$
- **Transition probability:** $a_{i,j,k} \equiv \Pr\{q_t = k|q_{t-1} = j, q_{t-2} = i, \Lambda\}$
- **Observation probability:** $b_k(\mathbf{x}) \equiv \Pr\{\mathbf{x}_t = \mathbf{x}|q_t = k, \Lambda\}$

Design an algorithm similar to the forward algorithm that is able to compute $\Pr\{X|\Lambda\}$ with a computational complexity of at most $\mathcal{O}\left\{TN^3\right\}$. Provide a proof that your algorithm has at most $\mathcal{O}\left\{TN^3\right\}$ complexity — this can be an informal proof.

Solution: Define $\alpha_t(i,j) = \Pr\{\mathbf{x}_1, \ldots, \mathbf{x}_t, q_{t-1} = i, q_t = j|\Lambda\}$. Compute it as

- **Initialize**:
$$\alpha_2(i,j) = \pi_{ij} b_i(\mathbf{x}_1) b_j(\mathbf{x}_2), \quad 1 \le i,j \le N$$
- **Iterate:**
$$\alpha_t(j,k) = \sum_{i=1}^{N} \alpha_{t-1}(i,j) a_{ijk} b_k(\mathbf{x}_t), \quad 1 \le t \le T,\ 1 \le j,k \le N$$
- **Terminate:**
$$\Pr\{\mathbf{X}|\Lambda\} = \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_T(i,j)$$

The highest-complexity part of the algorithm is the iteration step, which requires:

- for each of T different time steps t,
- for each of N different values of j,
- for each of N different values of k,
- we must compute a summation with N terms,

hence it has $\mathcal{O}\left\{TN^3\right\}$ complexity.

Problem 4.2

Suppose you have a sequence of $T = 100$ consecutive observations, $\mathbf{X} = [x_1, \ldots, x_T]$. Suppose that the observations are discrete, $x_t \in \{1, \ldots, 20\}$. You have it on good information that these data can be modeled by an HMM with $N = 10$ states, whose parameters are

- **Initial state probability:** $\pi_i \equiv \Pr\{q_1 = i|\Lambda\}$
- **Transition probability:** $a_{ij} \equiv \Pr\{q_t = j|q_{t-1} = i, \Lambda\}$
- **Observation probability:** $b_j(x) \equiv \Pr\{x_t = x|q_t = j, \Lambda\}$

In terms of these model parameters, and in terms of the forward probabilities $\alpha_t(i)$ and backward probabilities $\beta_t(i)$ (for any values of i, j, t, x that are useful to you), what is $\Pr\{q_{17} = 7, x_{18} = 3|x_1, \ldots, x_{17}, x_{19}, \ldots, x_{100}, \Lambda\}$?

Solution: Conditional = joint / marginal. The joint probability is

$$\Pr\{q_{17} = 7, x_1, \ldots, x_{17}, x_{18} = 3, x_{19}, \ldots, x_{100}\} = \sum_{j=1}^{10} \alpha_{17}(7) a_{7j} b_j(3) \beta_{18}(j)$$

The marginal is

$$\Pr\{x_1, \ldots, x_{17}, x_{19}, \ldots, x_{100}\} = \sum_{i=1}^{10} \sum_{j=1}^{10} \sum_{k=1}^{20} \alpha_{17}(i) a_{ij} b_j(k) \beta_{18}(j)$$

So the conditional is

$$\Pr\{q_{17} = 7, x_{18} = 3|x_1, \ldots, x_{17}, x_{19}, \ldots, x_{100}\} = \frac{\sum_{j=1}^{10} \alpha_{17}(7) a_{7j} b_j(3) \beta_{18}(j)}{\sum_{i=1}^{10} \sum_{j=1}^{10} \sum_{k=1}^{20} \alpha_{17}(i) a_{ij} b_j(k) \beta_{18}(j)}$$

Problem 4.3

A **partially-observed Markov model** is a model in which some part of the state variable is observed, while other parts are not observed. For example, consider a model with 2 states in which q_1 is observed to be $q_1 = 1$, and q_3 is observed to be $q_3 = 2$, but q_2 is not observed. This model has no output vectors (no $\mathbf{x}$): your only observations are the two state IDs, q_1 and q_3. All parts of this problem are cumulative; in your answer to any part, you may use any assumptions that were specified in any previous part.

(a) Suppose that you have a transition probability matrix $\mathbf{A}$, whose $(i, j)^{\text{th}}$ element is

$$a_{ij} = \Pr\{q_t = j|q_{t-1} = i\}$$

Find a formula in terms of the elements of $\mathbf{A}$ for

$$\gamma_2(j) = \Pr\{q_2 = j|q_1 = 1, q_3 = 2, \mathbf{A}\}$$

Solution:

$$\gamma_2(j) = \frac{a_{1j} a_{j2}}{\sum_{i=1}^{2} a_{1i} a_{i2}}$$

(b) The expected log likelihood, can be defined as

$$Q(\mathbf{A}', \mathbf{A}) = E\left[\ln \Pr\{q_1 = 1, q_2 = j, q_3 = 2|\mathbf{A}'\} \mid q_1 = 1, q_3 = 2, \mathbf{A}\right]$$

Find a formula for $Q(\mathbf{A}', \mathbf{A})$ in terms of the elements of $\mathbf{A}$ and $\mathbf{A}'$, and/or in terms of $\gamma_2(j)$.

Solution:

$$Q(\mathbf{A}', \mathbf{A}) = \sum_{j=1}^{2} \gamma_2(j) \left(\ln a'_{1j} + \ln a'_{j2}\right)$$

(c) The Lagrangian method for optimization works as follows. Suppose we are trying to find values of a'_{ij} that maximize $Q(\mathbf{A}', \mathbf{A})$, subject to the constraint that

$$\sum_{j=1}^{2} a'_{ij} = 1$$

The Lagrangian method creates a Lagrangian function $L(\mathbf{A})$ by creating a "constraint term" $(1 - \sum_j a'_{ij})$ that must be zero if the constraint is satisfied, multiplying the constraint term by a "Lagrangian multiplier" λ_i, and then adding the result to $Q(\mathbf{A}', \mathbf{A})$, resulting in :

$$L(\mathbf{A}') = Q(\mathbf{A}', \mathbf{A}) + \sum_{i=1}^{2} \lambda_i \left(1 - \sum_{j=1}^{2} a'_{ij}\right)$$

In terms of the elements of $\mathbf{A}'$, $\gamma_2(j)$, and the Lagrangian multipliers λ_1 and λ_2, what are the values of $dL(\mathbf{A}')/da'_{ij}$ for each value of $i, j \in \{1, 2\}$?

Solution:

$$\frac{dL(\mathbf{A}')}{da'_{11}} = \frac{\gamma_2(1)}{a'_{11}} - \lambda_1$$
$$\frac{dL(\mathbf{A}')}{da'_{12}} = \frac{1}{a'_{12}} - \lambda_1$$
$$\frac{dL(\mathbf{A}')}{da'_{21}} = -\lambda_2$$
$$\frac{dL(\mathbf{A}')}{da'_{22}} = \frac{\gamma_2(2)}{a'_{22}} - \lambda_2$$

(d) Set $\frac{dL(\mathbf{A}')}{da'_{11}} = 0$ and $\frac{dL(\mathbf{A}')}{da'_{12}} = 0$. Doing so will give you the new model parameters, a'_{11} and a'_{12}, in terms of both $\gamma_2(j)$ and λ_i. Choose a value of λ_i so that $a'_{11} + a'_{12} = 1$.

Solution:

$$a'_{11} = \frac{\gamma_2(1)}{1 + \gamma_2(1)}$$
$$a'_{12} = \frac{1}{1 + \gamma_2(1)}$$

Note: you are not asked to solve for a'_{21} because there's a trick: dL/da'_{21} cannot be set to zero. As long as $\lambda_2 > 0$, $dL/da'_{21} < 0$. The conclusion is that you should make a'_{21} as small as possible, thus:

$$a'_{21} = 0$$
$$a'_{22} = 1$$