

Homework #2

HMMs, Pair HMMs

Due at the beginning of class on Tuesday, February 9

Collaboration is allowed in groups of at most four students, but you must submit separate writeups. Please write the names of all your collaborators on your solutions. You are not allowed to copy group work. If you are working alone, we will drop the problem with the lowest score. If you submit your solutions after the submission deadline, you must write the date and time of submission on your writeup. Under no circumstances will a homework be accepted more than three days after its due date.

1. Problem 1 (20 points) Markov Chain Review and End States

Let x be a sequence of length L , where the x_i denotes the state at the i th position. We assume that x_i is the outcome by drawing a random variable X_i from a finite alphabet \mathcal{X} with K values (or states), and the sequence is generated via a Markov Chain, i.e., $a_{x_i, x_{i+1}} = P(x_{i+1}|x_i, x_{i-1}, \dots, x_1) = P(x_{i+1}|x_i)$. Suppose that the initial probabilities are given, i.e., $P(x_1)$ are known for $x_1 \in \mathcal{X}$.

- (5 points) Express the marginal probability of observing x_i at the i th position, i.e., compute $P(x_i)$ in terms of the transitional probabilities and initial probabilities given above.
- (5 points) Show that the probabilities of observing all sequences of length L sums to 1, i.e., $\sum_x P(x) = 1$. (By showing this, you have shown that the joint distribution over all possible sequences is a proper probability distribution.)
- (5 points) Suppose now that the length of the sequence is defined by an end state α , and at each position, each x_i can transition to this end state with probability $P(\alpha|x_i) = \tau$ and the Markov properties still apply. Compute the sum of the probabilities over all sequence of length L . In other words, compute

$$\sum_{x_1} \sum_{\{x_2: x_2 \neq \alpha\}} \dots \sum_{\{x_L: x_L \neq \alpha\}} P(x_1, x_2, \dots, x_L, \alpha).$$

- (5 points) Adding the end states gives us the ability to model sequences of varying lengths with a proper distribution. Show that the sum of probabilities of all possible sequences of *any* length is 1.

2. Problem 2 (17 points) HMMs True or False

For each of the following sentences, say whether it is true or false and provide a short explanation (one sentence or so). Questions (a)-(d) are worth 2 points and the remaining are worth 3 points each.

- An edge from state s to state t in an HMM denotes the conditional probability of going to state s given that we are currently at state t .
- The weights of all incoming edges to a state of an HMM must sum to 1.

- c. Given a sequence $x = x_1 \dots x_N$ and a parse $\pi = \pi_1 \dots \pi_N$ of x by a first order HMM, we can compute the likelihood of the parse $P(x, \pi)$ as follows:

$$P(x, \pi) = P(\pi_1)P(\pi_2|\pi_1)P(x_2|x_1) \dots P(\pi_N|\pi_{N-1})P(x_N|x_{N-1})$$

- d. The Baum-Welch algorithm is a type of an Expectation Maximization algorithm and as such it is guaranteed to converge to the (globally) optimal solution.
- e. The most likely parse output from of the Viterbi algorithm can be used to evaluate the most likely state at position i given the sequence x .
- f. The probability in a cell of the Viterbi matrix is never less than the probability in the corresponding cell of the Forward matrix.
- g. Using an HMM to model a process that depends on several time steps changes the time complexity from $O(K^2N)$ to $O(DK^2N)$ for decoding or evaluation, where D is the length of the duration.

3. **Problem 3 (23 points) A simple HMM example**

Assume that we have a Hidden Markov Model (HMM) as shown in Fig.1. For $i = 1, 2, 3$, X_i denotes the states in i -th position, and Y_i represents the corresponding observations.

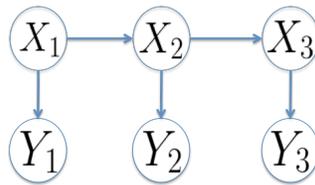


Figure 1: A simple HMM example.

- (a) (4 points) If there are k different states and a total of m different observations are possible. How many parameters are required to fully define this HMM. And also what conditional independences hold in this simple HMM? Please justify your answer.
- (b) (4 points) Now assume there are two possible states (S_1 and S_2) for X_i and similarly there are two possible values (0 and 1) for Y_i . You are given the following probabilities:

$$P(X_1 = S_1) = 0.99,$$

$$P(X_{i+1} = S_1|X_i = S_1) = 0.99, P(X_{i+1} = S_2|X_i = S_2) = 0.99,$$

$$P(Y_i = 0|X_i = S_1) = 0.8, P(Y_i = 0|X_i = S_2) = 0.9.$$

Now use both forward algorithm and backward algorithm to compute the probability that we observe the sequence $Y_1 = 0, Y_2 = 1, Y_3 = 0$. Please check whether both algorithms give the same results. (please provide detailed steps)

(c) (6 points) Using the forward-backward algorithm, report the most likely setting for the first three states. Show your work in detailed steps.

(d) (6 points) Use the Viterbi algorithm to compute the most likely setting for the first three states. Show your work in detailed steps.

(e) (3 points) Compare the results from (c) and (d). Does this make sense? Provide 1-2 sentences to justify for your answer.

4. Problem 4 (20 points) Initial Values in Training

(a) (5 points) Suppose we have a large number of sequences emitted by an HMM that has a particular transition probability $a_{kl} = 0$, for some k and l . Say that we now use these emitted sequences to train (using Baum-Welch) a new HMM with the same architecture, one that happens to start with $a_{kl} = 0$. Prove that the parameter a_{kl} will remain 0 after the training.

(b) (5 points) Assume we have an HMM to train, and we choose initial conditions such that two of the states, k and k' , are identical. That is, for all other states l , $a_{kl} = a_{k'l}$, $a_{lk} = a_{lk'}$, and for all characters b , $e_k(b) = e_{k'}(b)$. Prove or disprove that Baum-Welch will keep these two states identical. How about Viterbi training?

(c) (10 points) Initial parameters play an important role on the eventual result of HMM training. The Baum-Welch and Viterbi algorithms are not guaranteed to find the correct parameters or state sequence respectively. In this problem you are asked to demonstrate this with an example using Baum-Welch.

i- Create an HMM which is the true model of a random process, such as a coin toss with fair and loaded dice.

ii- Create a set of (one or more) training sequences. Their number and length is your choice, but they should satisfy one requirement: A_{kl} and $E_k(b)$, the counts of each transition and emission in all the sequences, should be exactly consistent with a_{kl} and $e_k(b)$, the true model parameters:

$$\forall k, l, b : a_{kl} = \frac{A_{kl}}{\sum_j A_{kj}} \text{ and } e_k(b) = \frac{E_k(b)}{\sum_x E_k(x)}$$

iii- Find a set of initial nonzero parameters for an HMM which has the same architecture as the true model, such that Baum-Welch converges to the true model. At least one of the parameters must be different from the corresponding parameter of the true model.

iv- Find another set of initial nonzero parameters, such that Baum-Welch converges to an incorrect model.

5. Problem 5 (20 points) Pair HMMs for Alignments Architecture

For the following problems, you may ignore start and end states, unless they are important for the function of the HMM. You do not need to give the exact probabilities for all edges and emissions, but you should explain how one could compute them from the parameters given.

- a. (4 points) What is the smallest pair HMM you can build that corresponds to running Needleman-Wunsch with linear (not affine) gap penalty (with parameters: match m , mismatch s , gap d).
- b. (4 points) Construct a pair HMM that performs regular Needleman-Wunsch alignment with linear (not affine) penalty, where 1) in any contiguous set of gaps, gaps in x must occur before gaps in y , and 2) all gaps in x are of length at least 2.
- c. (4 points) Construct a pair HMM for overlap-detection alignment, as described in lecture 2. Show the state diagram with transition probabilities and explain briefly.
- d. (4 points) Construct a pair HMM that performs regular Needleman-Wunsch alignment with affine gap penalty, but where the gaps can be of at most length L .
- e. (4 points) Construct a pair HMM that will perform alignment that is equivalent to a variant of Needleman-Wunsch with piecewise linear gaps, i.e., a gap of length k incurring penalty not $d + ek$, but $\min\{d_1 + e_1k, \dots, d_s + e_s k\}$, and which has at most $2s + 1$ states.