SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.007 Machine Learning, Fall 2021
Homework 4

Due Tuesday 30 November 2021, 5pm

This homework will be graded by Zhang Qi

In this homework, we would like to look at the Hidden Markov Model (HMM), one of the most influential models used for structured prediction in machine learning.

1. (10 pts) Assume that we have the following training data available for us to estimate the model parameters:

State sequence	Observation sequence
$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{X})$	$(\mathbf{b}, \mathbf{c}, \mathbf{a}, \mathbf{b})$
$(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$	$(\mathbf{a}, \mathbf{b}, \mathbf{c})$
$(\mathbf{Z}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}, \mathbf{Y})$	$(\mathbf{b}, \mathbf{c}, \mathbf{a}, \mathbf{b}, \mathbf{c})$
$(\mathbf{Z}, \mathbf{X}, \mathbf{Y})$	$(\mathbf{c}, \mathbf{b}, \mathbf{a})$

   Clearly state what are the parameters associated with the HMM. Under the maximum likelihood estimation (MLE), what would be the values for the optimal model parameters? Clearly show how each parameter is estimated exactly.

2. (10 pts) Now, consider during the evaluation phase, you are given the following new observation sequence. Using the parameters you just estimated from the data, find the most probable state sequence using the Viterbi algorithm discussed in class. Clearly present the steps that lead to your final answer.

State sequence	Observation sequence
$(?, ?)$	$(\mathbf{b}, \mathbf{c})$

3. (10 pts) The Viterbi algorithm discussed in class can be used for computing the single most probable state sequence for a new observation sequence. Specifically, in the Viterbi algorithm we are interested in finding the optimal sequence using the following formula:

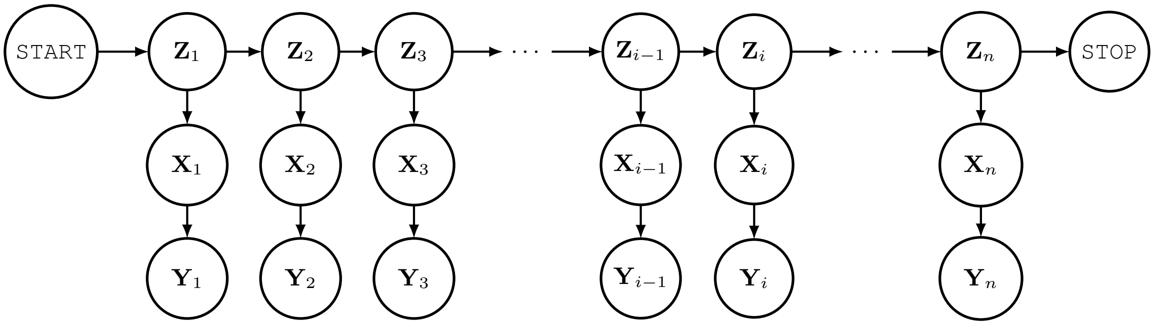
$$(s_1^*, s_2^*) = \arg\max_{s_1, s_2} P(s_1, s_2 | o_1 = \mathbf{b}, o_2 = \mathbf{c})$$

   However, an alternative method to estimate the state sequence is finding the optimal state of index i by solving the optimization problem

$$s_i^* = \arg\max_{s_i} P(s_i | o_1 = \mathbf{b}, o_2 = \mathbf{c})$$

   where $P(s_i | o_1 = \mathbf{b}, o_2 = \mathbf{c})$ is the marginal distribution and $i \in \{1, 2\}$. Please use the forward and backward scores for HMM discussed in class to calculate the marginal distribution $P(s_i | o_1 = \mathbf{b}, o_2 = \mathbf{c})$. Clearly present the steps that lead to your final answer.

4. (20 pts) Now consider a slightly different graphical model which extends the HMM (see below). For each state ($\mathbf{Z}$), there is now an observation pair ($\mathbf{X}$, $\mathbf{Y}$), where $\mathbf{Y}$ sequence is generated from the $\mathbf{X}$ sequence.



Assume you are given a large collection of observation pair sequence, and a predefined set of possible states, you would like to estimate the most probable state sequence for each observation pair sequence using an EM algorithm similar to the dynamic programming algorithm discussed in class. Clearly define the forward and backward scores in a way analogous to those defined for HMM, and explain what they mean. Give algorithms for computing the forward and backward scores. Analyze the complexity associated with your algorithms.