SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

01.112 Machine Learning, Fall 2018
Homework 4

Due Friday 23 Nov 2018, 5pm

This homework will be graded by Thilini Cooray

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

   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.

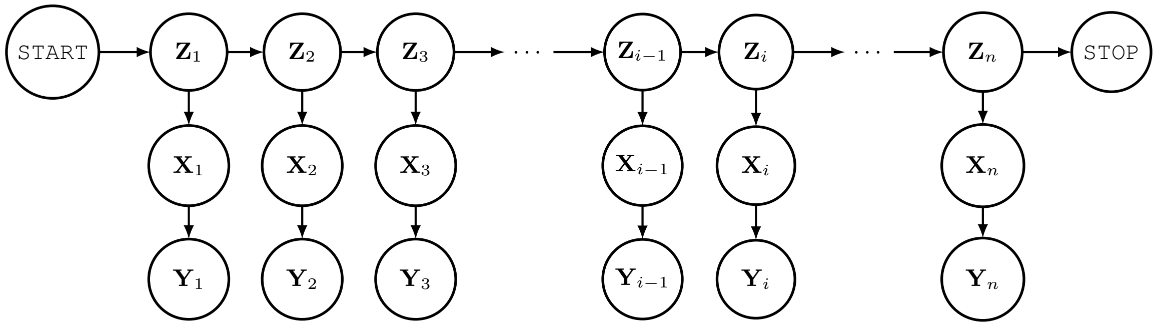
2. (10 pts) Now, consider during the evaluation phase, you are given the following new observation sequence. Using the parameters you just estimated from the data, find the most probable state sequence using the Viterbi algorithm discussed in class. Clearly present the steps that lead to your final answer.

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

3. (20 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:

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

   However, sometimes we are interested in finding the k most probable state sequences for a given observation sequence. This is sometimes called top-k decoding. Clearly describe how to modify the Viterbi algorithm to support top-k decoding.



4. (20 pts) Now consider a slightly different graphical model which extends the HMM (see above). 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 that we discussed in class, and explain what they mean. Give algorithms for computing the forward and backward scores. Analyze the time complexity associated with your algorithms.