SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

50.007 Machine Learning, Fall 2015
Homework 4

Due Monday 23 Nov 2015, 5pm

This homework will be graded by Dinh Quang Thinh.
Please submit the hard copy of your solutions to 2.716-S2.

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

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
$(?, ?)$	$(\mathbf{a}, \mathbf{d})$

3. (20 pts) The HMM discussed in class makes a simple first-order assumption, where the next state only depends on the previous state in the generative process. However, it is possible to extend the model discussed in class to have second-order dependencies. In other words, the HMM can be parameterised in the following way:

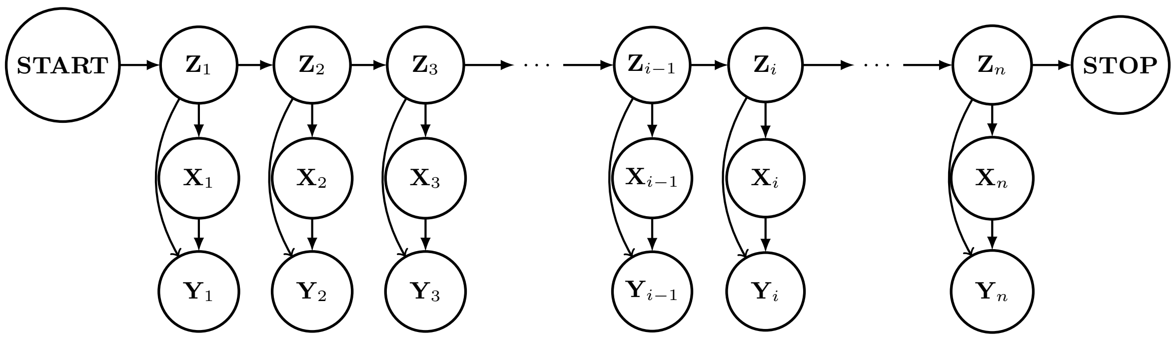
$$p(x_1, \ldots, x_n, y_1, y_2, \ldots, y_n) = \prod_{i=1}^{n+1} p(y_i | y_{i-2}, y_{i-1}) \cdot \prod_{i=1}^{n} p(x_i | y_i)$$

where we define $y_{-1} = y_0 = \texttt{START}$ and $y_{n+1} = \texttt{STOP}$.

In other words, the transition probabilities are changed from $p(y_i|y_{i-1})$ to $p(y_i|y_{i-2}, y_{i-1})$ now. Describe the Viterbi algorithm used for decoding such a second-order HMM model. In other words, describe the dynamic programming algorithm that computes the following efficiently for such an HMM:

$$(y_1^*, y_2^*, \ldots, y_n^*) = \underset{y_1, y_2, \ldots, y_n}{\arg\max}\, p(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n)$$

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 both the $\mathbf{X}$ sequence and $\mathbf{Z}$ sequence.



Assume you are given a large collection of observation pair sequences, and a predefined set of possible states $\{0, 1, \ldots, N-1, N\}$, where $0 = \textbf{START}$ and $N = \textbf{STOP}$. You would like to estimate the most probable state sequence for each observation pair sequence using an algorithm similar to the dynamic programming algorithm discussed in class. Clearly define the forward and backward scores in a way analogous to HMM. Give algorithms for computing the forward and backward scores. Analyze the time complexity associated with your algorithms (for an observation pair sequence of length n).