	Barry Tee Wei Cong 1001549	Date No.
1		# (u;v)
	Parameter 1: Transition Probabilities, an, = cour	ant (u)
	u\√ X Y Z	ment STOP with him had
	START 2/4 0/4 2/4	0/4
	X 0/5 2/5 2/5	1/5 2 1 4 1 4
	Y 1/5 0/5 1/5	3/5 ()
	7 2/5 3/5 0/5	CAN 10/54 AVEN 4
	must 1	(u→0)
) - A	Parameter 2: Emission Probabilities, by (0) = county	nt (u)
	Color wort (by 1-	- T. W - S(D 1)7
ESFA.	12: 10 UNO passe u la ma la 16: 2 and come	not be yet property
	X 1/5 3/5 1/5	110 70 400
C.Y		THE TANK OF THE PARTY OF THE PA
	2 1/5 2/5 1000 2/5	TOTAL BURE OF THE STORE
2.	· Base Case	
6.4	$\pi$ (0, start) = 1	
	I Possession Co.	orsheri os i
71.0	· Recursive Case	TO A KIND OF SELECTION OF THE SELECTION
	$T(1,X) = T(0,START) \cdot ASTART, X \cdot b_X(a) = 1 \cdot \frac{3}{2}$	
	$\pi(1, Y) = \pi(0, START) \cdot \alpha_{START, Y} \cdot b_{Y}(\alpha) = 1 \cdot \frac{1}{2}$	
	$T(1,7) = T(0,START) \cdot a_{START} \cdot b_{2}(a) = 1 \cdot a_{START}$	
	TO CITE - TO CONTINE ! . MS[PIK] , Z DE (W)	7 - 10
	$\Pi(2,X) = \max_{u \in T} \left\{ \Pi(1,u) \cdot a_{u,X} \cdot b_{X}(a) \right\}$	
	$= \max \left[ \frac{1}{16} \cdot \frac{1}{5} \cdot \frac{1}{5$	\frac{1}{7} = \frac{350}{2}
	$\Pi(Z,Y) = \max_{u \in T} \left\{ \Pi(I,u) \cdot \alpha_{u,Y} \cdot b_{X}(a) \right\}$	
	= max [ to · \forall · \fo	\frac{2}{5} \cdot = \frac{6}{350} = \frac{3}{120}
	$T(2,2) = \max_{u \in T} \left\{ T(1,u) \cdot \alpha_{u,2} \cdot b_{2}(\alpha) \right\}$	5 ) 250 123
	= max { to · = · + · to · = · + · to · = ·	上 } = ====
+	- IIIVIN ( IV 3 3 , V 3 3 , IO 3	2 1 720
	$\pi(3,STOP) = \max_{v \in T} \{\pi(2,v) \cdot \alpha_{v,STOP}\}$	
	Maloroll Act correctly and all the	
	$= \max \left[ \frac{2}{250} \cdot \frac{1}{5}, \frac{3}{125} \cdot \frac{3}{5}, \frac{2}{250} \cdot \frac{9}{5} \right]$	

		TAYAT 1	06/00	2.815-6		tre A .	-
٠.	γj≠u	(30) 194 4	A V	route defilier	ARCHART NOTHER	it the yest pive	mer
-	A lind the		V 4-	· ·			
	Modified Viter				A	Wid II	
1				4/0		O#403	
-				in the free Y		<u> </u>	
-	•			Ivin · bu(xj)			
$\dashv$	→ Move	torward recu	arshely wh	m j=2,,n	tor all u.	4	
+	Also	10 EV	, troves	Was to the	Kala Call r Mila	in 7 Livin	
1	TI: 11	= MOX TT (i	(-1 u)	0 b (v.	holostañ aszin <b>)</b>	CA - TA ADDAGE	(A. HD 4
+							
+					n for all u		W-101
1	-						
+	7 move	TOWNYN PECH	rsively wh	ne = 2,, 1	n for all u	except tov	W= VON
+	3. TT (n+1 )	3(01)-	11 (N , N) - 1	M w ,STOP	<u> </u>	<u> </u>	
+							
						- 1966 - 1966	106.
	1				1 = (78	972 . GO T	
					1 = (78	977 . CO <sup>TT</sup>	
	4. Baurtrackiv	NOJ			1 s (T8)	T (0, STB 10, STB	5.00% ·
14	4. Bautrachiv	nax {TII	1,u)·a <sub>n,s</sub>	отор <u>}</u>	) = (T)		
	4. Bayetracking  y * = are yn-1 = are	na o max { TT (n	1,u)·α <sub>u,s</sub>	5TOP }	I = (TA		
	$y_{n-1}^* = \frac{\alpha r \alpha}{\alpha r \alpha}$	O max {The	2	(6) VA . V 19	STREET . CTERTS	,¢) / ≃ (X.	
	$y_{n-1}^* = \frac{\alpha r \alpha}{\alpha r \alpha}$	O max {The	2	(6) VA . V 19		,¢) / ≃ (X.	
	$y_{n-1}^* = \frac{\alpha r \alpha}{\alpha r \alpha}$	O max {The	2	(6) VA . V 19	STREET . CTERTS	,¢) / ≃ (X.	
	$y_{n-1}^* = \frac{\alpha r_n}{4}$ $y_{n-1}^* = \frac{\alpha r_n}{4}$ $y_n^* = \frac{\alpha r_n}{4}$	Omax {TII	i, u) · au	(a) vd . v . v . v . v . v . v . v . v . v	STREET . CTERTS	,¢) / ≃ (X.	
	$y_{n-1}^* = \frac{\alpha r_n}{4}$ $y_{n-1}^* = \frac{\alpha r_n}{4}$ $y_n^* = \frac{\alpha r_n}{4}$	O max {The	i, u) · au	(a) vd . v . v . v . v . v . v . v . v . v	or all u Texce	,¢) / ≃ (X.	
	$y_{n-1}^* = \frac{\alpha r_n}{4}$ $y_{n-1}^* = \frac{\alpha r_n}{4}$ $y_n^* = \frac{\alpha r_n}{4}$	Omax {TII	i, u) · au	(a) vd . v . v . v . v . v . v . v . v . v	or all u Texce	1) Xully = (X XXully = (Y XXully = (X XXully = (X	MAN.
	$y_{n-1}^* = \frac{\alpha r}{\alpha r}$ $y_{n-1}^* = \frac{\alpha r}{\alpha r}$ $y_1^* = \frac{\alpha r}{\alpha r}$	O max { The	j, u) · Qu	1 (a) xd	or all w Texay	(c) / = (X, 10) / = (Y, 2+	MAN.
	$y_{n-1}^* = \frac{\alpha r}{\alpha r}$ $y_{n-1}^* = \frac{\alpha r}{\alpha r}$ $y_1^* = \frac{\alpha r}{\alpha r}$	Omax {TII	j, u) · Qu	(a) xd . x 72 (b) xd . x 4 . 0 1 (a) xd . xd . x 4 . 0 1 (a) xd . xd . x 5 . 0	x.wo (n.1)	1) Your = (X 1) Your = (X 1) Your = (X 1) Your = (X	T(2.)
	$y_{n-1}^* = arr$ $y_{n-1}^* = arr$ $y_1^* = arr$	O max { TI (n)	, u) · Ou	1 (n) xd	TRANS (DILL)	(2) T = (X, (0) T = (Y, (1)	T(2.)
	$y_{n-1}^* = arr$ $y_{n-1}^* = arr$ $y_1^* = arr$	O max { The	, u) · Ou	1 (n) xd	TRANS (DIS)	(2) T = (X, (0) T = (Y, (1)	T(2.)
	$y_{n-1}^* = arr$ $y_{n-1}^* = arr$ $y_1^* = arr$	O max { TI (n)	, u) · Ou	(A) xd . x . y . y . y . y . y . y . y . y . y	YINR IN IN	CIFE = (X.   O)     = (Y.   O)     = (Y.   O)     = (Y.   O)	11 (2) T
	$y_{n-1}^* = arr$ $y_{n-1}^* = arr$ $y_1^* = arr$	O max { TI (n)	, u) · Ou	(A) x d	YINR (NI)	XAM = (X   XAM = (X	11 (2) T
	$y_{n-1}^* = arr$ $y_{n-1}^* = arr$ $y_1^* = arr$	O max { TI (n)	, u) · Ou	(A) x d	YINR IN IN	XAM = (X   XAM = (X	11 (2) T

	Date No.
4.	Assume we have a set of possible states {0,1,, N-1, N} where 0 = START and N=STOP
	P(x <sub>1</sub> ,, X <sub>i-1</sub> , y <sub>i</sub> ,, y <sub>i-1</sub> , Z <sub>i</sub> =u, x <sub>i</sub> ,, x <sub>n</sub> , y <sub>i</sub> ,, y <sub>n</sub> ;θ) =P(x <sub>1</sub> ,, X <sub>i-1</sub> , y <sub>i</sub> ,, y <sub>i-1</sub> , Z <sub>i</sub> =u;θ)·P(x <sub>i</sub> ,, x <sub>n</sub> , y <sub>i</sub> ,, y <sub>n</sub>   Z <sub>i</sub> =u;θ) = α <sub>n</sub> (i)·β <sub>n</sub> (i)
A 18185	
Visit (in the second	Forward Algorithm
	· Base Case
	$\alpha_{n}(1) = \alpha_{\text{START}, u}$ , $\forall u \in \{1,, N-1\}$
	· Recursive Case for $j=2,,n$
	$d_{u}(i+1) = \sum_{i} \alpha_{v}(i) \cdot \alpha_{v,u} \cdot b_{v}(x_{i}) \times c_{v}(y_{i})$ where $c_{v}(y) = P(y v)$
	Backward Algorithm
	Base Case
	$B_{n}(n) = \alpha_{u,stop} \cdot b_{n}(x_{n}) \times C_{n}(y_{n})  \forall n \in \{1,,N-1\}$
	Recursive (ase for n-1,, 1
	$\beta_n(n) = \sum_i \beta_r(i+1) \cdot b_n(x_i) \cdot a_{u,v} \times C_n(y_i)$
	The standards in the
	The length of the sentence is n.  For each position / layer, there are T forward & T backward terms.
- 1	Total: $O(n) = O(n T^2)$
	ISTAL TO COLY - COLY -
45.	



## 01.112 Machine Learning, Spring 2018 Homework 4

Due 4 April 2018, 5pm

## This homework will be graded by Allan Jie

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.

**Question 1.** Assume that we have the following training data available for us to estimate the model parameters:

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

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. (10 points)

**Question 2.** 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. (10 points)

$$\begin{array}{c|c} \textbf{State sequence} & \textbf{Observation sequence} \\ \hline (?,?) & (\mathbf{a},\mathbf{a}) \\ \end{array}$$

**Question 3.** The Viterbi algorithm discussed in class is used for finding the optimal y sequence based on the following:

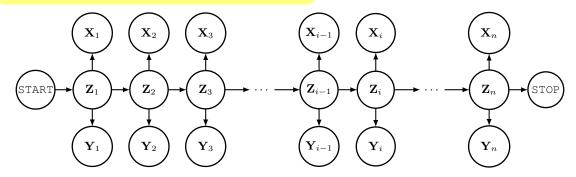
$$y_1^*, \dots, y_n^* = \arg\max_{y_1, \dots, y_n} p(x_1, \dots, x_n, y_1, \dots, y_n)$$

Now, consider the problem of part-of-speech tagging. Sometimes we have some prior knowledge about certain tags for certain observations. For example, assume the observation  $x_i$ ="the", we are almost certain that it is not a verb (i.e., we believe  $y_i \neq V$ ). In this case, we would like to do the decoding in the following way, where we would like to incorporate the prior knowledge  $y_i \neq V$  (and find optimal values for all other  $y_k$  in the sequence, where  $k = 1, \ldots, n, k \neq i$ ):

$$y_1^*, \dots, y_{i-1}^*, y_{i+1}^*, \dots, y_n^* = \underset{y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n}{\operatorname{arg max}} p(x_1, \dots, x_n, y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_n | y_i \neq V)$$

Clearly describe how to modify the Viterbi algorithm to perform such a new decoding task. (10 points)

**Question 4.** 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  $\mathbf{X}$  sequence is generated from the  $\mathbf{Z}$  sequence and  $\mathbf{Y}$  sequence is also generated from the  $\mathbf{Z}$  sequence.



Assume you are given a large collection of observation pair sequences, and a predefined set of possible states  $\{0, 1, ..., N-1, N\}$ , where  $0 = \mathbf{START}$  and  $N = \mathbf{STOP}$ . 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). (10 points)