HW NB

UMass CS 585
10/13/16
Viterbi algorithm

① Log-Prob of a doc, given class

$$P(\vec{w}_d \mid y) = \prod_{t=1}^{T} P(w_t \mid y)$$

$$\log(\text{———}) = \sum_t \log P(w_t \mid y)$$

Count same word multiple times?

② Derive under NB $P(\vec{w}_d) = \sum_{y \in \{NEG, POS\}} P(\vec{w} \mid y)\, P(y)$

$$= \sum_y \left[\prod_t P(w_t \mid y)\right] P(y)$$

HMM: $P(\vec{w}) = \sum_{\vec{y}} \underbrace{P(\vec{w}, \vec{y})}$

w_t

$w_{d,t}$

$\sum_{\vec{y}} \rightarrow K^T$ poss. seqs

$$= P(\vec{y})\, P(\vec{w} \mid \vec{y})$$

$$= \underline{\prod_{t=1}^{T} P(y_t \mid y_{t-1})\, P(w_t \mid y_t)}$$

What good is an HMM?

Model: $P(\vec{w},\vec{y}) = \prod_t \underbrace{P(y_t \mid y_{t-1})} \, \underbrace{P(w_t \mid y_t)}$

① Likelihood (LM) calc $P(\vec{w}) = \underbrace{\sum_{\vec{y}}}_{\cancel{\text{brute sum}}} P(\vec{w},\vec{y})$

Algo: Forward Algo

(NN, VB, NN)

② Tagging/Pred/Decoding: calc $\arg\max_{\vec{y}} \underbrace{P(\vec{y} \mid \vec{w})} = \dfrac{P(\vec{w},\vec{y})}{\cancel{P(\vec{w})}}$

Algo: Viterbi

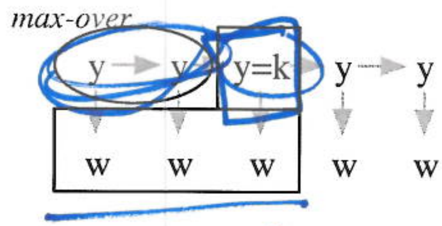
NN?
VB?

$\arg\max_{y_1, y_2, y_3} P(w_1 \mid y_1)\, P(y_2 \mid y_1)\, P(w_2 \mid y_2)\, P(y_3 \mid y_2)\, P(w_3 \mid y_3)$

Viterbi algorithm

Declaratively:

$$V_t[k] = \max_{y_1 \cdots y_{t-1}} P(y_t = k, \quad y_1..y_{t-1}, w_1..w_t)$$



$V_3[k]$

Algorithm, for each *t=1..N*,

For each k

$$V_t[k] := \max_{j=1..K} \left(V_{t-1}[j] \; P_{trans}(k \mid j) \; P_{emit}(w_t \mid k) \right)$$

$B_t[k] := \operatorname{argmax}_j (\text{———})$

$(\log V_t[k]) := \max(\log V_{t-1}(j) + \log P(\cdot) + \log P(\cdot$

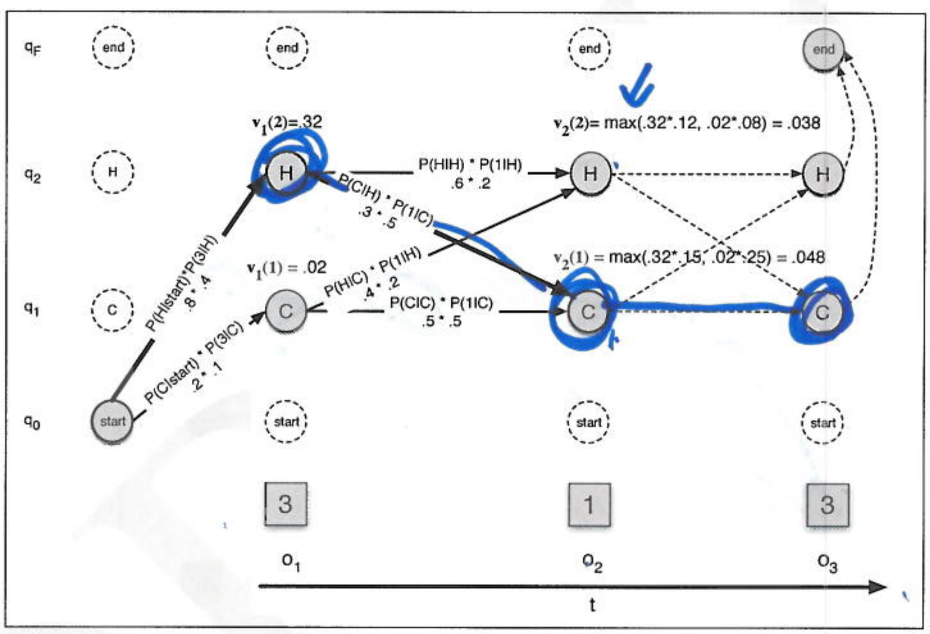
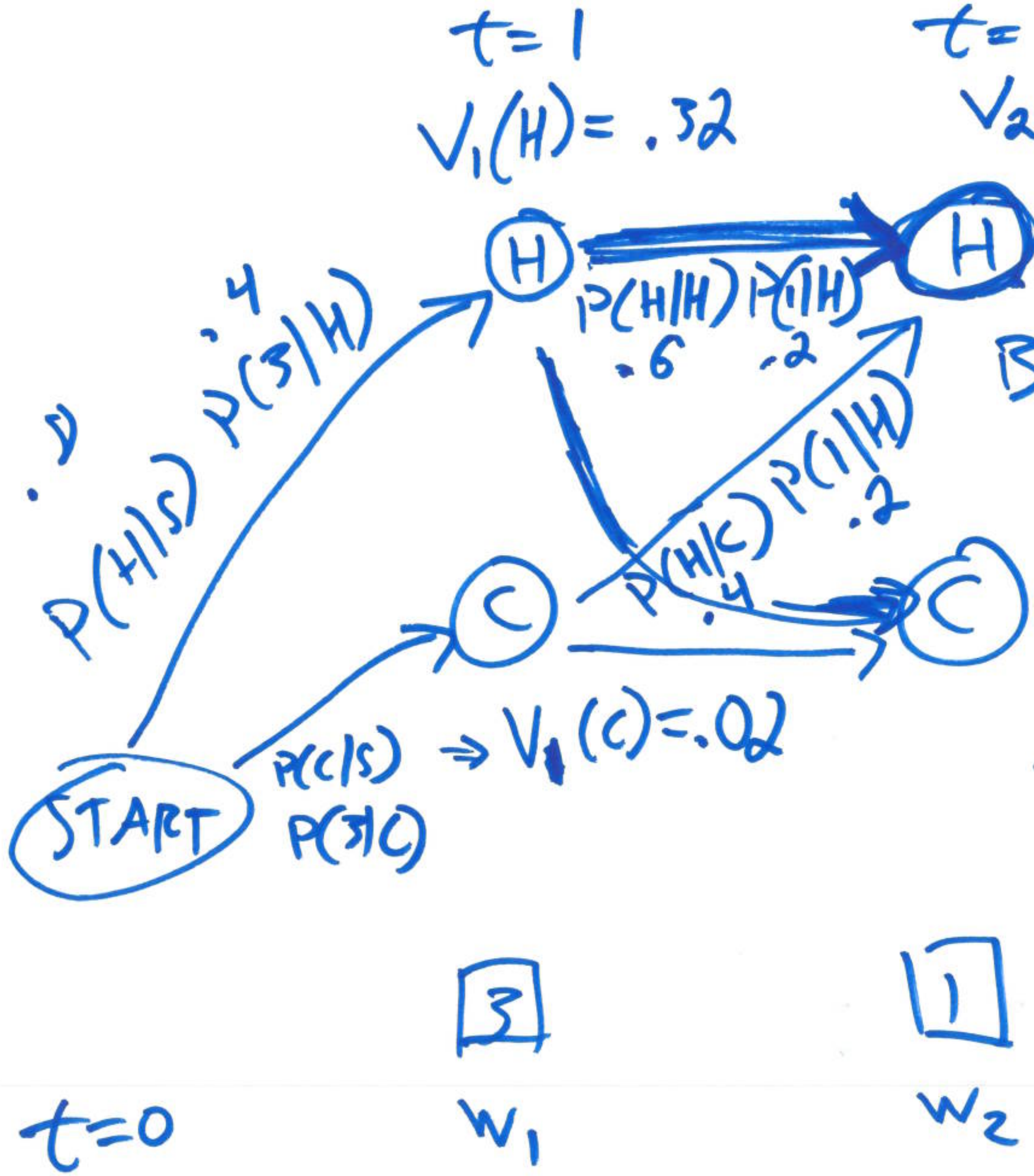


Figure 7.10 The Viterbi trellis for computing the best path through the hidden state space for the ice-cream eating events *3 1 3*. Hidden states are in circles, observations in squares. White (unfilled) circles indicate illegal transitions. The figure shows the computation of $v_t(j)$ for two states at two time steps. The computation in each cell follows Eq. 7.19: $v_t(j) = \max_{1 \leq i \leq N-1} v_{t-1}(i)\, a_{ij}\, b_j(o_t)$. The resulting probability expressed in each cell is Eq. 7.18: $v_t(j) = P(q_0, q_1, \ldots, q_{t-1}, o_1, o_2, \ldots, o_t, q_t = j \mid \lambda)$.



$$V_2(H) = \max(.32 \cdot .6 \cdot .2,\ .02 \cdot .4 \cdot .2) = .038$$

$$B_2(H) = \operatorname{argmax}(\text{—}, \text{—}) = H$$

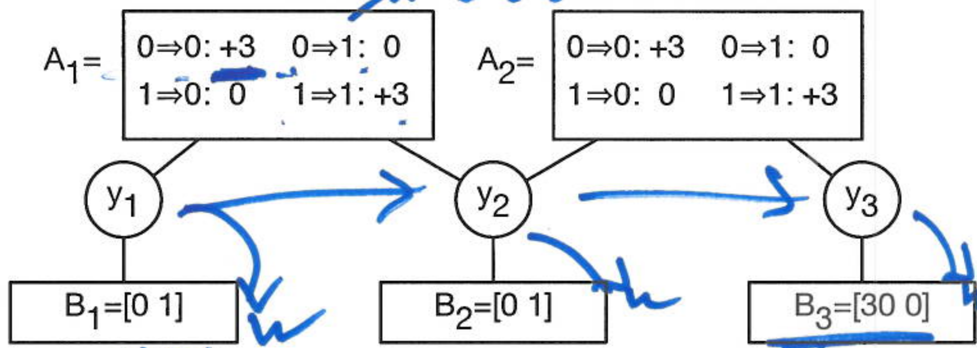
$$V_2(C) = \text{—} \cdot \text{—} \cdot \text{—}$$

$$B_2(C) = H$$

CS585 Viterbi in-class exercise
Oct 13 2016

Name: ______________________________

$A(0,0)=3$ $A(0,1)=0$



$B_1(0)=0$
$B_1(1)=1$

$B_3(0)=30$
$B_3(1)=0$

Sticky-favoring ~~CRF~~ HMM over vocab {0,1}. Factor scores are in log-scale additive form

$$G(y1, y2, y3) = A(y1,y2) + A(y2,y3) + B1(y1) + B2(y2) + B3(y3)$$

Most probable sequence: G( ) =

$\underset{y_1 y_2 y_3}{\arg\max}\ G(y_1 y_2 y_3)$

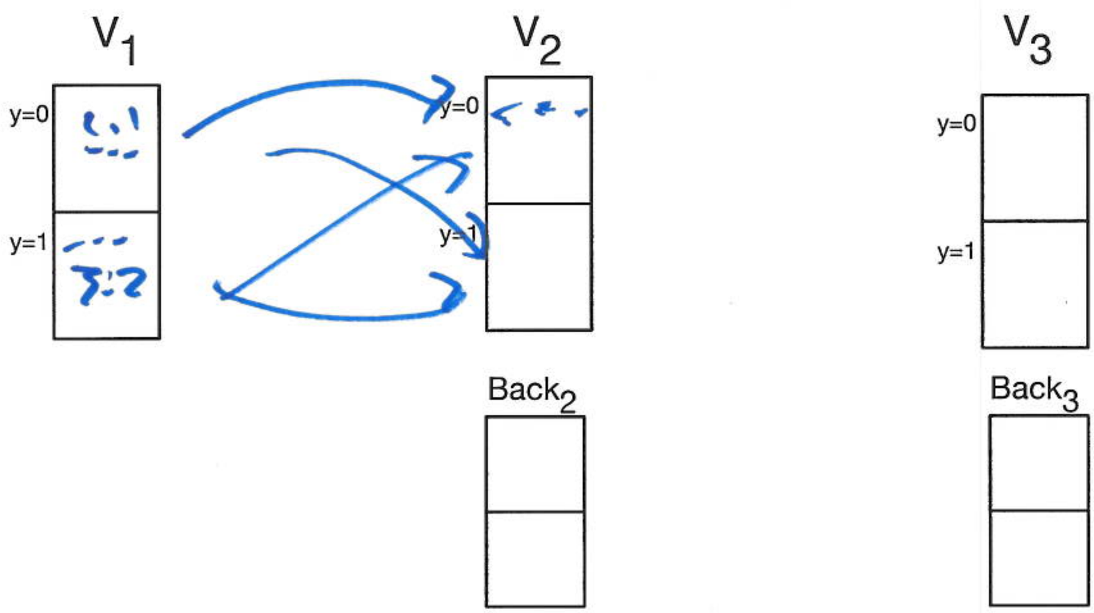
$\log P(k|j)$ $\log P(w_t|k)$

Viterbi in additive form....	For t=1..T, For k in {0,1}, $V_t[k] := \max_j \left(V_{t-1}[j] + A(j,k) + B_t(k)\right)$ $Back_t[k] := \arg\max_j (\ldots)$

For t=1, assume A_0(anything)=0 and V_0[anything]=0

Final backtrace step: take best-scoring from last V_T, follow the backpointers all the way back

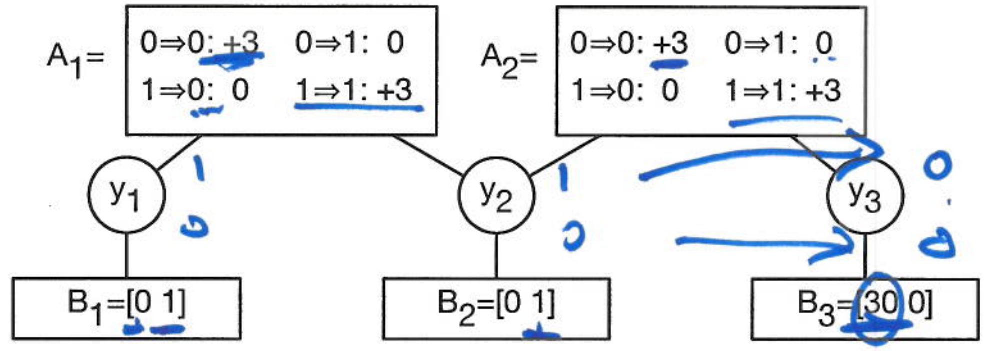
Run Viterbi and fill out the trellis with arcs like in the textbook's HMM example.



Solution y* = (, ,)

CS585 Viterbi in-class exercise
Oct 13 2016

Name: ______________________________



Sticky-favoring CRF over vocab {0,1}. Factor scores are in log-scale additive form

$\arg\max_{y_1 y_2 y_3}$ G(y1, y2, y3) = A(y1,y2) + A(y2,y3) + B1(y1) + B2(y2) + B3(y3)

1 1 0
vs 0 0 0

Most probable sequence: G(0, 0, 0) = 30 + 3 + 3 = 36

Viterbi in additive form....

For t=1..T,

For k in {0,1},

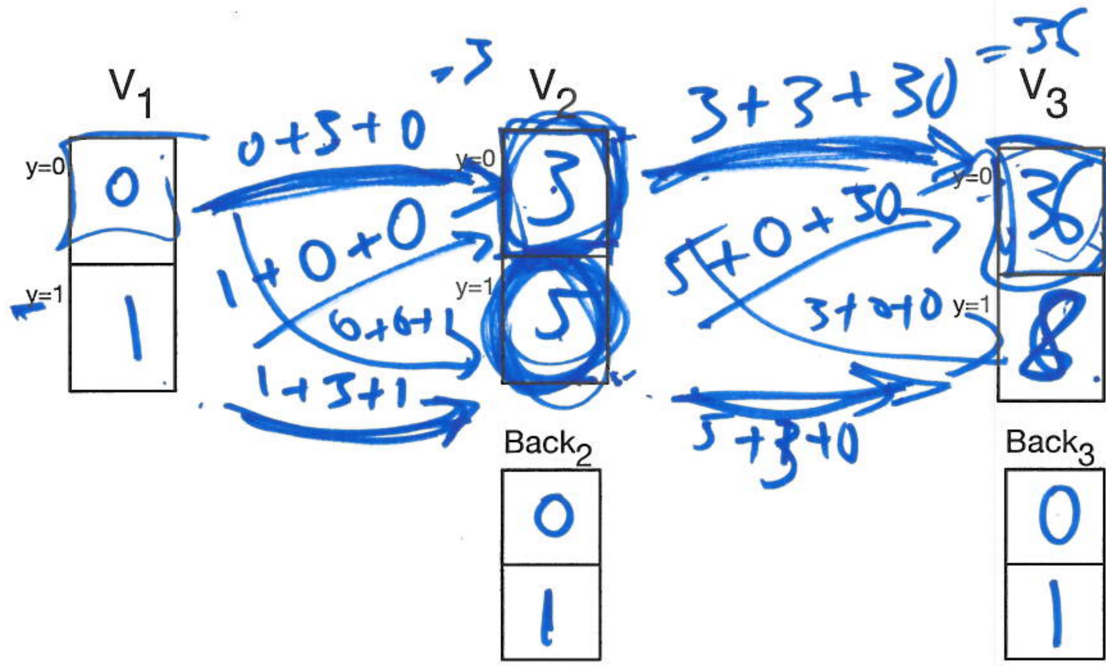
$$V_t[k] := \max_j \left(V_{t-1}[j] + A(j,k) + B_t(k)\right)$$

$$Back_t[k] := \arg\max_j (\ldots)$$

For t=1, assume A_0(anything)=0 and V_0[anything]=0

Final backtrace step: take best-scoring from last V_T, follow the backpointers all the way back

Run Viterbi and fill out the trellis with arcs like in the textbook's HMM example.



Solution y* = (0, 0, 0)