CS 378 Lecture 11: Viterbi

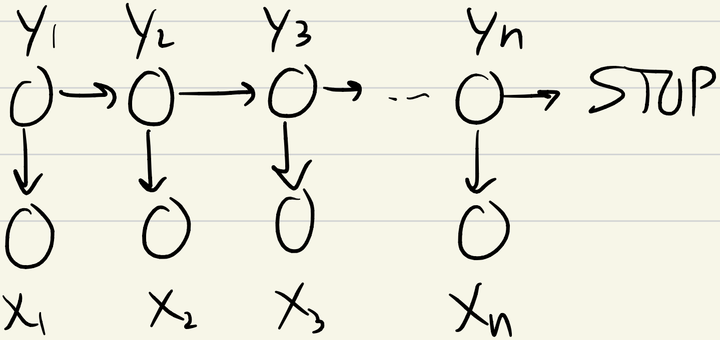
Announcements

- Midterm in 2 weeks
- A3 due in 9 days
- Survey
  - More exs
  - More code discussion
- Final proj

Recap HMMs

Tags $y_i \in \mathcal{T}$ words $x_i \in \mathcal{V}$

$$P(\bar{y}, \bar{x}) = P(y_1)\, P(x_1 \mid y_1)\, P(y_2 \mid y_1)\, P(x_2 \mid y_2)\, P(y_3 \mid y_2)\, P(x_3 \mid y_3) \cdots P(\text{STOP} \mid y_n)$$



Parameters: Initial $P(y_1)$ $|\mathcal{T}|$

Transitions $P(y_i | y_{i-1})$ $|\mathcal{T}-1| \times |\mathcal{T}|$

Emissions $P(x_i | y_i)$ $|V| \times |\mathcal{T}|$

Training: count + normalize

Inference: Viterbi

$$\underset{\bar{y}}{\operatorname{argmax}}\ P(\bar{y} | \bar{x})$$

$$= \underset{\bar{y}}{\operatorname{argmax}}\ P(\bar{y}, \bar{x})$$

$$= \underset{\bar{y}}{\operatorname{argmax}}\ \log P(\bar{y}, \bar{x})$$

Ex Log probabilities (approx.)

$S =$

N	−1
V	−1

$T =$

	N	V	STOP
N	−2	−1 ☆	−1
V	−1	−1	−2

$E =$

	they	fish	can
N	−1	−1	−3
V	−3	−1	−1

they can fish: what is the most likely tag sequence?

8 choices

NNN (STOP)

NNV

NVN

⋮

Can compute scores explicitly:

−1   −1☆   −1   −2

N V V

−1 they   −1 can   −1 fish

score: −8

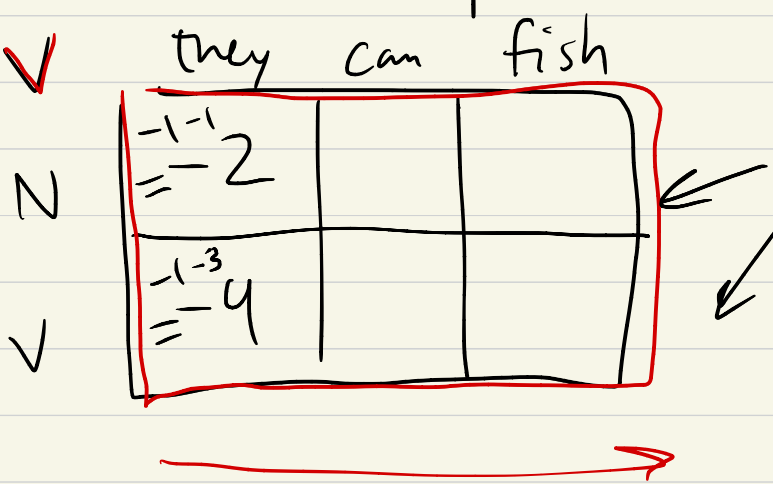
NVN: ☆ score −7 [best]

Viterbi Dynamic Program

Define $v_i(\tilde{y})$ — i is an index in the sent $\{1, 2, 3\}$

$n \times |\mathcal{T}|$ matrix — $\tilde{y} \in \mathcal{T}$

$v_i(\tilde{y})$ = log prob of the best tag sequence ending in $\tilde{y}$ at step i



at the end: which is higher?

Initial:

$$v_1(\tilde{y}) = \underbrace{\log P(x_1 | \tilde{y})}_{\text{emissions}} + \underbrace{\log P(\tilde{y})}_{\text{initial}}$$

Recurrent: compute v_i using v_{i-1}

$$v_i(\tilde{y}) = \log P(x_i | \tilde{y}) + \max_{\tilde{y}_{prev}} \left[\log P(\tilde{y} | \tilde{y}_{prev}) + v_{i-1}(\tilde{y}_{prev}) \right]$$

Viterbi: for $i = 1 \ldots n$

for $\tilde{y} \in \mathcal{T}$

Compute $v_i(\tilde{y})$

$$v_{n+1}(\text{STOP}) = \max_{\tilde{y}_{prev}} \log P(\text{STOP} | \tilde{y}_{prev}) + v_n(\tilde{y}_{prev})$$

Track "backpointers" to get sequence

$S =$

N	-1
V	-1

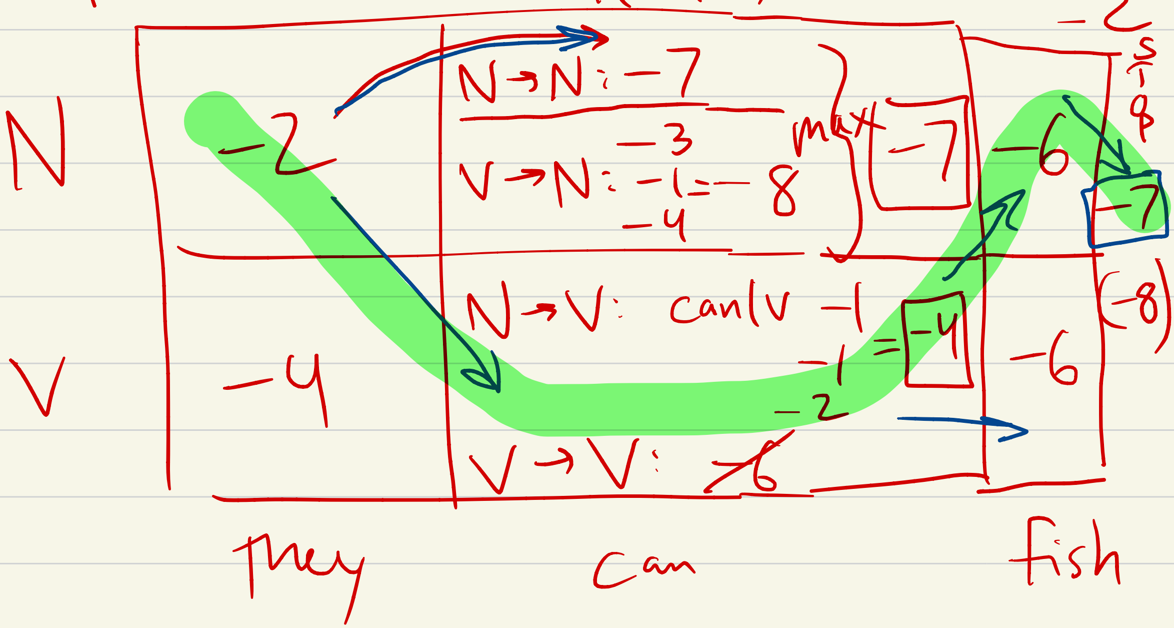
$T =$

	N	V	STOP
N	-2	-1 *	-1
V	-1	-1	-2

$E =$

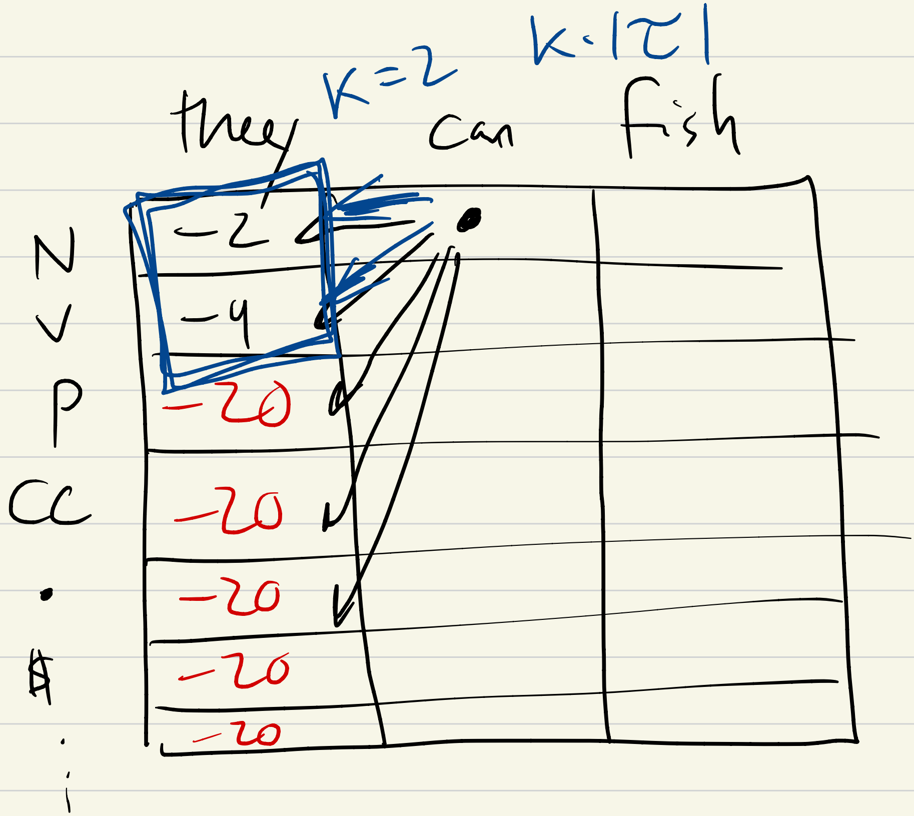
	they	fish	can
N	-1	-1	-3
V	-3	-1	-1

$V_2(N)$
part 1: N→N: emission $P(can|N) = -3$
$+ P(N|N): -2 + V_1(N)$
-2



Beam Search $|\mathcal{T}|$, n

Viterbi runtime: $O(n|\mathcal{T}|^2)$



Beam search: keep top k scores in each column priority queue

N: -2
V: -4
J: -19

CC: -20
• : -20

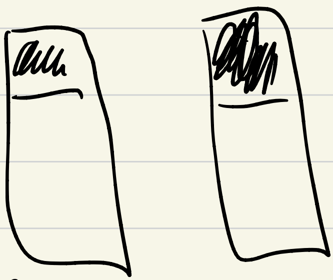
kicked out

(overhead from priority queue)

$$O(n|T|^2) \rightarrow O(n|T|k \log|T|)$$

data structure

"2D" beaming



prevs transitions