CS 378 Lecture 10

Today

- Review HMMs
- Viterbi Algorithm
- Beam search
- If time: revisit POS tagging

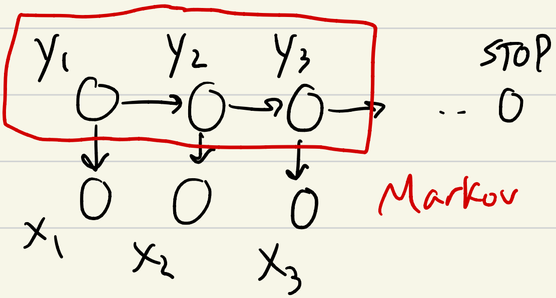
Announcements

- A2 due tonight
- A3 out tonight

Recap HMMs: sequence model

tag $\bar{y}$, $y_i \in \mathcal{T}$ words $\bar{x}$, $x_i \in V$

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



Markov process

Params: start probs $P(y_1)$
transitions $P(y_i | y_{i-1})$
emissions $P(x_i | y_i)$

Inference in HMMs

Given $\bar{x}$. Want to return most likely $\bar{y}$

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

① argmax over an exponentially large space $|\mathcal{Y}|^n$

② we don't have $P(\bar{y}|\bar{x})$

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

$$P(\bar{x}) = \sum_{\bar{y}} P(\bar{y},\bar{x})$$ prob. of the sent.

const. w.r.t. $\bar{y}$

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

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

$$\underset{\tilde{y}_1, \dots, \tilde{y}_n}{\operatorname{argmax}} \log P(\tilde{y}_1) + \log P(x_1 | \tilde{y}_1) + \log P(\tilde{y}_2 | \tilde{y}_1) + \log P(x_2 | \tilde{y}_2) + \dots$$

<u>Ex</u> log probs for HMM

$$S = \begin{array}{l} N: -1 \\ V: -1 \end{array} \qquad T = \begin{array}{c|ccc} & N & V & STOP \\ \hline N & -2 & -1 & -1 \\ V & -1 & -1 & -2 \end{array}$$

$$E = \begin{array}{c|ccc} & \text{they} & \text{can} & \text{fish} \\ \hline N & -1 & -3 & -1 \\ V & -3 & -1 & -1 \end{array}$$

$P(x_i | y_i)$

they can fish

What is the most likely tag seq?

N V V   N N V   ..

N V N   N N N

-1 -1 -1 -2

init t t t

N V V STOP total log prob: -8

they can fish

e e e

-1 -1 -1

-1 -1 -1 -1 log prob = -7 ✓

N V N STOP

th can fi

-1 -1 -1

<u>Viterbi Algorithm</u> Dynamic programming

Define $V_i(\tilde{y})$ $n \times |\mathcal{T}|$ matrix

sent len

log prob of the best tag seq ending in $\tilde{y}$ at step i

	they	can	fish
N	-2		
V	-4		

V_i

Initial: $v_1(\tilde{y}) = \log P(x_1 | \tilde{y}) + \log P(\tilde{y})$

Recurrent: We can figure out scores for step i using scores from step $i-1$

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

<u>Viterbi</u> for $i = 1 \ldots n$

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

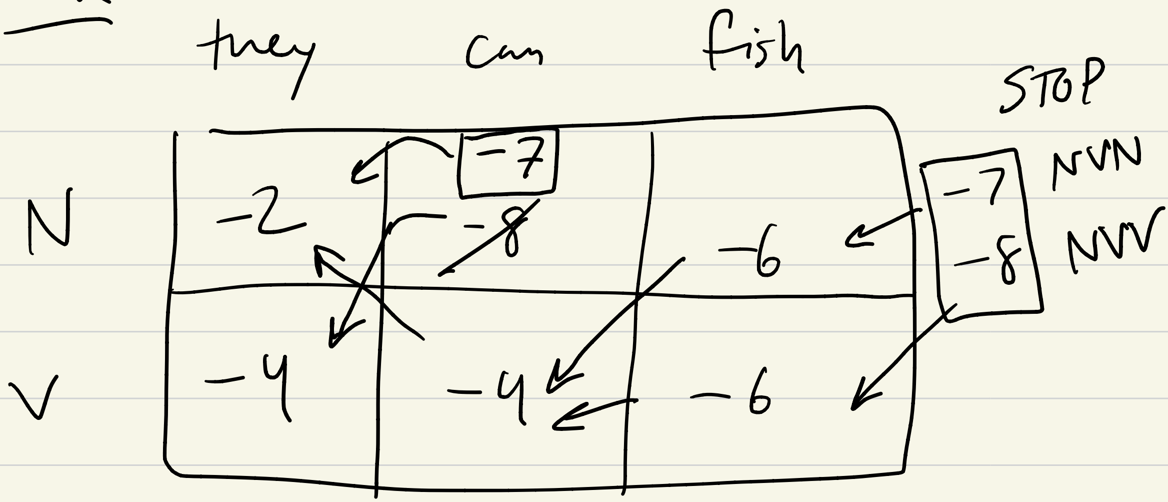
Compute $v_i(\tilde{y})$ as above

Add STOP transition to end

Compute $v_{n+1}(\text{STOP}) = \max_{\bar{y}} \log P(\bar{y}, \bar{x})$

Track "backpointers" to reconstruct sequence

Ex



$$S = \begin{array}{l} N: -1 \\ V: -1 \end{array} \qquad T = \begin{array}{c|ccc} & N & V & STOP \\ \hline N & -2 & -1 & -1 \\ V & -1 & -1 & -2 \end{array}$$

$$E = \begin{array}{c|ccc} & \text{they} & \text{can} & \text{fish} \\ \hline N & -1 & -3 & -1 \\ V & -3 & -1 & -1 \end{array}$$

$$V_2(N) = \log P(\text{can} \mid N) \rightarrow \boxed{-3}$$

$$+ \max_{\tilde{y}_{prev}} \begin{cases} N: \log P(N \mid N) + V_1(N) & \underbrace{}_{-4} \quad \underbrace{}_{-2} \quad \overset{-2}{} \\ V: \log P(N \mid V) + V_1(V) & \underbrace{}_{-5} \quad \underbrace{}_{-1} \quad \underbrace{}_{-4} \end{cases}$$

Getting 2nd best

New example

① N V N

② V V N

Recovering K-best list is hard!
n-best

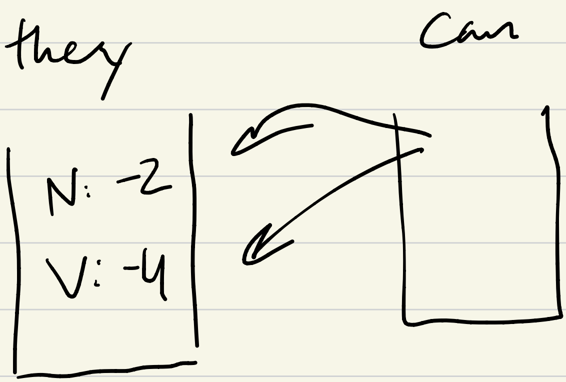
Beam Search

Viterbi: $O(n|T|^2)$ slow if $|T|$ is big

Most transitions suck

most words can't take most tags

Beam search: only keep top K scoring states in each column V_i of Viterbi chart



only 2 prev options, not $|T|$

beam size = 2
everything not in top 2 gets kicked out

can: N
V
P
CC
.
$
;

$O(|T|)$ tags to check, but only K for y_{prev}

$$O(n|T|^2)$$

$$\Rightarrow O(n|T|K)$$

(but approximate!)

"sort of"

- log K from dealing with priority queue