Section A: HMMs

Monday, November 7, 2022 10:07 AM

3 Problems for an HMM

Evaluation: $p(\vec{x}) = \sum_{\vec{y} \in \mathcal{Y}_{\vec{x}}} p(\vec{x}, \vec{y})$

$$p(\vec{x},\vec{y}) = p(y_1)\left[\prod_{t=1}^{T} p(x_t|y_t)\right]\left[\prod_{t=2}^{T} p(y_t|y_{t-1})\right]$$

For $|\vec{y}| = T$, and $y_t \in \{1, \dots, K\}$ there are K^T possible labelings $\vec{y} \in \mathcal{Y}_{\vec{x}}$

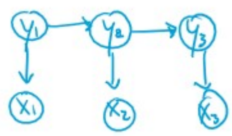
$|\mathcal{Y}_{\vec{x}}| = K^T$

Viterbi Decoding: $\hat{y} = \underset{\vec{y} \in \mathcal{Y}_{\vec{x}}}{\text{argmax}}\ p(\vec{y}|\vec{x})$

$$\frac{p(\vec{x},\vec{y})}{p(\vec{x})}$$

Marginals: $p(y_t = k|\vec{x}) = \sum_{\substack{\vec{y} \in \mathcal{Y}_{\vec{x}} \\ \text{s.t. } y_t = k}} p(\vec{y}|\vec{x})$

Ex:



$y_t \in \{I, H, V\}$

$x_t \in \{C, N\}$

joint

$$p(x_1, x_2, x_3, y_1, y_2, y_3) = p(x_1|y_1)\, p(x_2|y_2)\, p(x_3|y_3)\, p(y_3|y_2)\, p(y_2|y_1)\, p(y_1)$$

① Evaluation

$$p(x_1, x_2, x_3) = \sum_{y_1} \sum_{y_2} \sum_{y_3} p(\vec{x}, \vec{y})$$

② Viterbi Dec.

$$\hat{y} = \underset{y_1, y_2, y_3}{\text{argmax}}\ p(y_1, y_2, y_3 | x_1, x_2, x_3)$$

③ Marginals

$$p(y_2 = k | x_1, x_2, x_3) = \sum_{y_1} \sum_{y_3} p(y_1, y_2, y_3 | x_1, x_2, x_3)$$

Public Service Announcement

BEFORE

implementing the F-B algo

FIRST

implement the Brute Force algorithm

Brute Force for Evaluation

```
def eval(x):
    p_x = 0
```

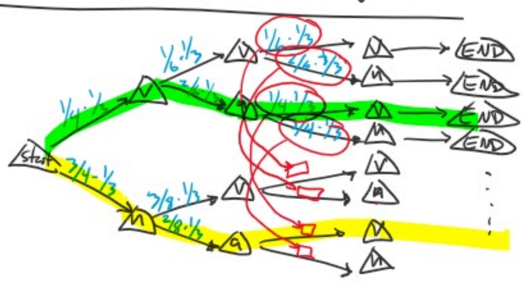
Terrible Idea

```
def eval(x):
    p_x = 0
    for y in all_y(x):
        p_x += joint(x,y)
    return p_x
```

Terrible Idea

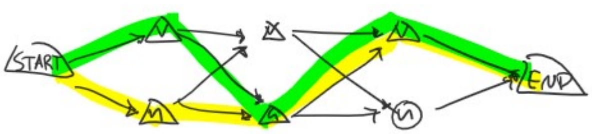
Good Idea

F-B Search Space aka. $\mathcal{Y}_{\vec{x}}$



START find prefered tags END

Key Idea: the current tag holds all the information for the scoring of the next tag (Markov assumption)



Trained Model w/ parameters A and B

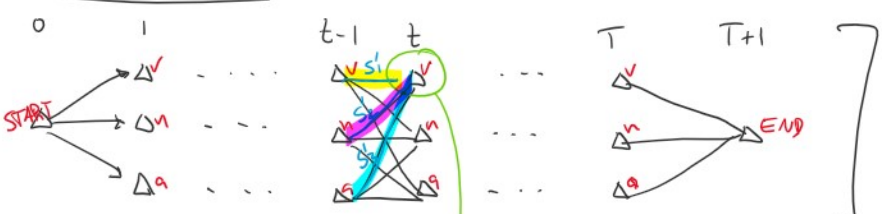
A

y_{t-1}	y_t	$p(y_t \mid y_{t-1})$
v	v	1/6
v	a	2/6
v	n	2/6
v	END	1/6
n	v	3/8
n	a	2/8
n	END	3/8
a	v	1/4
a	n	3/4
START	v	1/4
START	n	3/4

B

y_t	x_t	$p(x_t \mid y_t)$
v	find	1/3
v	prefered	1/3
v	tags	1/3
n	find	1/3
n	prered	1/3
n	tags	1/3
⋮	⋮	⋮

F-B Algo



trellis

$\alpha_t(v) = \alpha_{t-1}(v)\, s'_1 + \alpha_{t-1}(n)\, s'_2 + \alpha_{t-1}(a)\, s'_3$

= total weight of path prefixes entering node (t, v)

forward (edge weights):

for $t = 1, \dots, T$:

  for $k = 1, \dots, K$:

   $\alpha_t(k) = \sum_{j=1}^{K} \alpha_{t-1}(j)\, s_{kjt}$

For HMM case:

$s_{kjt} = p(y_t = k \mid y_{t-1} = j)\, p(x_t = w \mid y_t = k)$

$$s_{kjt} = p(y_t = k \mid y_{t-1} = j)\, p(x_t = w \mid y_t = k)$$
$$= A_{j,k}\, B_{k,w}$$

<u>backward (edge weights):</u>

same also, but on the mirror image of the trellis

$\boxed{071}$

$$p(x_1 \mid y, x_2) = \frac{p(x_1, y, x_2)}{p(y, x_2)}$$
$$= \frac{p(x_1 \mid y)\, p(x_2 \mid y)\, p(y)}{p(y, x_2)}$$
$$= \frac{p(x_1 \mid y)\, \cancel{p(x_2, y)}}{\cancel{p(x_2, y)}}$$
$$= p(x_1 \mid y) \Rightarrow x_1 \perp\!\!\!\perp x_2 \mid y$$

$$p(A \mid B, C) = p(A \mid C) \Rightarrow A \perp\!\!\!\perp B \mid C$$
$$p(A, B \mid C) = p(A \mid C)\, p(B \mid C) \Rightarrow$$

$$\boxed{p(x_1 \mid y_1, x_2) = p(x_1 \mid y)}$$ b/c N.B. assumption

$\forall m, n$

$\forall m,n \quad x_m \perp\!\!\!\perp x_n \mid y$

MAP Estimation

$$J(\theta) = \underbrace{\left[\sum_{i=1}^{N} \log p(y^{(i)} \mid x^{(i)}; \theta) \right]}_{\text{likelihood}} + \underbrace{\log p(\theta)}_{\text{prior}}$$

$= \quad \text{''} \quad + \underline{\underline{\|\theta\|_2}}$

L2 reg $\iff$ Gaussian prior

L1 reg $\iff$ Laplace prior