Section A

Monday, March 27, 2023 9:26 AM

3 Problems for an HMM

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

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

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

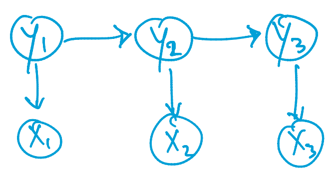
$|\mathcal{Y}| = K^T$

Viterbi Decoding: $\hat{y} = \arg\max_{\vec{y}\in\mathcal{Y}} \boxed{p(\vec{y}|\vec{x})}$

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

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

Ex:



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

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

① **Evaluation:**

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

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

② **Viterbi Decoding:**

$\hat{y} = \arg\max_{y_1,y_2,y_3} p(y_1,y_2,y_3|x_1,x_2,x_3)$

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

implement the

FIRST

Implmt the Brute Force Algo

Brute Force for Evaluation

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

Terrible Idea

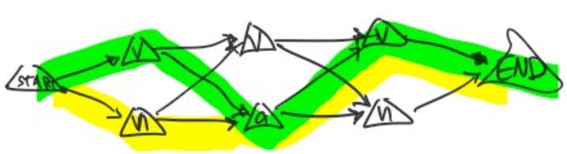
Good Idea

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

START find preferred tags END

Why? 1st Order Markov Assumption

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



Given: Trained Model Parameters A and B

B

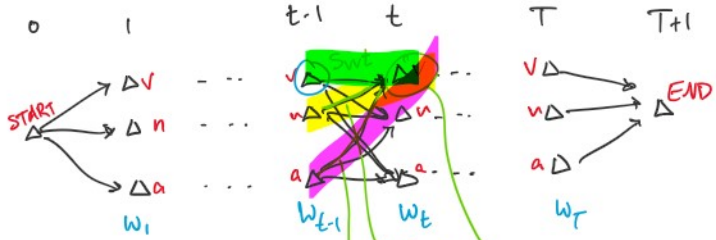
y_{t-1}	y_t	$p(y_t \mid y_{t-1})$
v	v	1/6
v	a	2/6
v	n	2/6
v	END	1/6
n	v	3/8
n	a	2/8
n	END	3/8
a	v	1/4
a	n	3/4
START	v	1/4
START	n	3/4

A

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	prefered	1/3
n	tags	1/3
⋮	⋮	

F-B Algo

F-B Algo



trellis

forward (edge weights):

$\alpha_t(v) = \alpha_{t-1}(v)\, s_{vvt} + \alpha_{t-1}(n)\, s_{vnt} + \alpha_{t-1}(a)\, s_{vat}$

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

for $t = 1 \ldots T$:

for $k = 1 \ldots K$:

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

For HMM Case:

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

backward (edge weights):

Same algo, but on the mirror image of the trellis

OH

$p(D|\theta) = \quad p(x_1, \ldots, x_N | \theta)$

$p(\_ \mid \_)$

likelihood

$\ell_{MLE}(\theta; D) = \ell^D_{MLE}(\theta) = \underline{p(x_1, \ldots, x_N | \theta)}$

posterior

$\ell(\ \mid\ )$

$\ell_{MAP}(\theta; D) = \ell^D_{MAP}(\theta) = \underline{p(\theta | x_1, \ldots, x_N)} \propto$ likelihood $\times$ prior

$p(\theta)$

$$\ell_{MAP}(\vec{\theta}) = \log \left(\prod_{i=1}^{N} p(y^{(i)} | x^{(i)}, \theta) \right) + \log p(\vec{\theta})$$

$p(\theta_m) \sim \text{Gaussian}(\mu = 0, \sigma^2 = 1/4$

$p(\theta) = \prod_{m=1}^{M} p(\theta_m)$

$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x \quad)^2}{2}\right)$

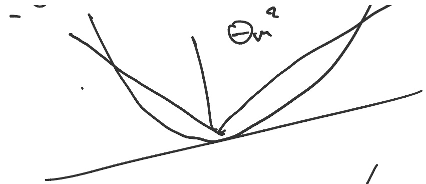
$\log p(\theta) = \left(-\sum_{m=1}^{M} \theta_m^2\right) + M \cdot \log(1/2 \ldots)$

$p(\theta_m) = \frac{1}{1/2\sqrt{2\pi}} \exp\left(-\frac{\theta_m^2}{\cancel{2}}\right)$

L2

$\log p(\theta_m) = \log(\sqrt{2\pi}\,1/2) +$

$\log p(\theta_m) = \log\left(2\pi\right)^{1/2} +$



L2 $\quad \sum_{m=1}^{M} \theta_m^2$

L1 $\quad \sum_{m=1}^{M} |\theta_m|$