Ex 5 Solution

Dienstag, 16. Mai 2023 22:35

T1 answer provided

T2 template: $\langle y_t, y_{t-1} \rangle$

$$\phi_1^{trans} = [\![y_t = 1 , y_{t-1} = 1]\!]$$
$$\phi_2^{trans} = [\![y_t = 2 , y_{t-1} = 1]\!]$$
$$\phi_3^{trans} = [\![y_t = 1 , y_{t-1} = 2]\!]$$
$$\phi_3^{trans} = [\![y_t = 2 , y_{t-1} = 2]\!]$$

template: $\langle y_t, y_{t-1} \rangle$

$$\phi_1^{obs} = [\![y_t = 1 , x_t = \text{"We"}]\!]$$
$$\phi_2^{obs} = [\![y_t = 1 , x_t = \text{"are"}]\!]$$
$$\phi_3^{obs} = [\![y_t = 1 , x_t = \text{"clever"}]\!]$$

$$\phi_4^{obs} = [\![y_t = 2 , x_t = \text{"We"}]\!]$$
$$\phi_5^{obs} = [\![y_t = 2 , x_t = \text{"are"}]\!]$$
$$\phi_6^{obs} = [\![y_t = 2 , x_t = \text{"clever"}]\!]$$

T3

use that for $a \geq 0$:

$$\max(a+b, a+c) = a + \max(b, c)$$

let $a(y_{t-1}, y_t) = \sum_{k=1}^{d_{trans}} w_k^{trans} \phi_k^{trans}(y_t, y_{t-1})$

$b(y_t, x, t) = \sum^{d_{obs}} w^{obs} \phi^{obs} (\ldots)$

$$b(y_t, x, t) = \sum_{k=1}^{d_{obs}} \omega_k^{obs} \phi_k^{obs}(x, y_t)$$

Searching for

$$\max_{y_1,\dots,y_T} \sum_{t=2}^{T} a(y_{t-1}, y_t) + \sum_{t=1}^{T} b(y_t, x, t) + \underbrace{\log Z(x)}$$

= const wrt. y (we will ignore this)

$$\leadsto \max_{y_1 \dots, y_T} \left[\sum_{t=2}^{T} a(y_{t-1}, y_t) + \sum_{t=2}^{T} b(y_t, x, t) + \underbrace{b(y_1, x, 1)}_{=: \delta_1(y_1)} \right]$$

$$= \max_{y_2, \dots, y_T} \left[\sum_{t=3}^{T} a(y_{t-1}, y_t) + \sum_{t=3}^{T} b(y_t, x, t) + \underbrace{\max_{y_1} \left(\delta_1(y_1) + a(y_1, y_2) \right) + b(y_2, x, t)}_{=: \delta_2(y_2)} \right]$$

$=: \delta_2(y_2)$ ← not a function of y_1 anymore

$\Psi_2(y_2)$: store argmax y_1 per y_2

$$= \max_{y_3, \dots, y_T} \left[\sum_{t=4}^{T} a(y_{t-1}, y_t) + \sum_{t=4}^{T} b(y_t, x, t) + \underbrace{\max_{y_2} \left(\delta_2(y_2) + a(y_2, y_3) \right) + b(y_3, x, t)}_{=: \delta_3(y_3)} \right]$$

$=: \delta_3(y_3)$ ← not a function of y_2 anymore

$\Psi_3(y_3)$: store argmax y_2 per y_3

$= \dots$

[$\delta_i(\cdot)$ and $\psi_i(\cdot)$ are discrete functions, with domain and range Σ each,

with $\forall t: y_t \in \Sigma$]

Now let $\Sigma = \{1, \dots, N\}$, "without loss of generality":

for $i := 1, \dots, N$

$\delta_1(i) = b(i, x, 1)$

$\psi_1(i) = -1$ ("not defined" / we will not use this)

for $t = 2, \dots T$

for $j = 1, \dots, N$

$$\delta_t(j) = \max_i \left(\delta_{t-1}(i) + a(i,j)\right) + b(j, x, t)$$

$$\psi_t(j) = \text{argmax}_i \left(\delta_{t-1}(i) + a(i,j)\right) + \underbrace{b(j, x, t)}_{\text{constant wrt } i \,/\, \text{can be ignored}}$$

$\hat{y}_T = \text{argmax}_i\, \delta_T(i)$

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

$\hat{y}_t = \psi_{t+1}(\hat{y}_{t+1})$