Yiwen Zhang

YZ3310

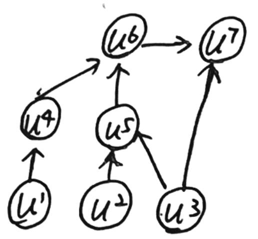
homework 4.

part#1 :

a_2 = SHIFT , a_3 = Left-Arc^{nsubj} , a_4 = SHIFT , a_5 = Right-Arc^{dobj} , a_6 = Right-Arc^{root}

part #2 :

Question 1 :



Question 2 :

$u^4 = 3 \times u^1 = 3$

$u^5 = u^2 \times u^3 = 2 \times 3 = 6$

$u^6 = (u^4)^2 + (u^5)^2 = 3^2 + 6^2 = 9 + 36 = 45$

$u^7 = u^3 + 10 \cdot u^6 = 3 + 10 \times 45 = 453$

Question 3 :

$$J^{3\to5}(u^2, u^3) = \frac{\partial f^5(u^2, u^3)}{\partial u^3} = u^2$$

$$J^{4\to6}(u^4, u^5) = \frac{\partial f^6(u^4, u^5)}{\partial u^4} = 2u^4$$

$$J^{5\to6}(u^4, u^5) = \frac{\partial f^6(u^4, u^5)}{\partial u^5} = 2u^5$$

$$J^{3\to7}(u^3, u^6) = \frac{\partial f^7(u^3, u^6)}{\partial u^3} = 1$$

$$J^{6\to7}(u^3, u^6) = \frac{\partial f^7(u^3, u^6)}{\partial u^6} = 10$$

Question 4 : There are 2 paths from u^3 to u^7, they are $u^3 \to u^5 \to u^6 \to u^7$ and $u^3 \to u^7$.

2,
$$\frac{\partial u^7}{\partial u^3}\Big|^{h^7}_{u^1, u^2, u^3} = J^{3\to7}(u^3, u^6) + J^{6\to7}(u^3, u^6) \times J^{5\to6}(u^4, u^5) \times J^{3\to5}(u^2, u^3)$$

$$= 1 + 10 \times 2u^5 \times u^2$$

$$= 1 + 10 \times 2 \times 6 \times 2$$

$$= 241$$

part3

Question 5 :

$\therefore x = \begin{bmatrix} 10 \\ -20 \end{bmatrix}$

$h_1 = RELU(W_1 \cdot x + b_1) = RELU(\langle 1, 1 \rangle \cdot \begin{bmatrix} 10 \\ -20 \end{bmatrix} + 0) = 0$

$h_2 = RELU(W_2 \cdot x + b_2) = RELU(\langle 1, 0 \rangle \cdot \begin{bmatrix} 10 \\ -20 \end{bmatrix} + 0) = 10$

$h_3 = RELU(W_3 \cdot x + b_3) = RELU(\langle 1, -1 \rangle \cdot \begin{bmatrix} 10 \\ -20 \end{bmatrix} + 0) = 30$

part #4

Question 6:

① $x_1^i = \text{the}$, $x_2^i = a$, $x_3^i = \text{this}$

$\because x_1^i, x_2^i, x_3^i \in \{\text{the}, a, \text{this}\}$

$\therefore u = x_1' + x_2' + x_3' = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

② $x_1^i = \text{the}$, $x_2^i = a$, $x_3^i = \text{mouse}$

$x_1^i, x_2^i \in \{\text{the}, a, \text{this}\}$ $\quad x_3^i \in \{\text{dog}, \text{cat}, \text{mouse}\}$

$\therefore u = x_1' + x_2' + x_3' = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

③ $x_1^i = \text{the}$, $x_2^i = \text{dog}$, $x_3^i = \text{mouse}$

$\because x_1^i \in \{\text{the}, a, \text{this}\}$ $\quad x_2^i, x_3^i \in \{\text{dog}, \text{cat}, \text{mouse}\}$

$\therefore u = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

④ $x_1^i = \text{cat}$, $x_2^i = \text{dog}$, $x_3^i = \text{mouse}$

$x_1^i, x_2^i, x_3^i \in \{\text{dog}, \text{cat}, \text{mouse}\}$

$\therefore u = x_1' + x_2' + x_3' = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

Question 7

We want to have $h>0$, ∵ $h>0 \Rightarrow W \cdot u + b > 0 \Rightarrow W \cdot u - 2 > 0 \Rightarrow \langle 1,0 \rangle \cdot u > 2$

∴ we only need to check what value for u could make $\langle 1,0 \rangle \cdot u > 2$ to be true.

u ~~have~~ has 4 possible values under these conditions:

① all x_1^i, x_2^i and x_3^i are from {the, a, this}, then $u = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$

② 2 of x_1^i, x_2^i and x_3^i from {the, a, this} and 1 from {dog, cat, mouse}, then $u = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$

③ 1 of x_1^i, x_2^i and x_3^i from {the, a, this} and 2 from {dog, cat, mouse}, then $u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

④ all x_1^i, x_2^i and x_3^i from {dog, cat, mouse}, then $u = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

so we check all possible u one by one:

1) if $u = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$, $\langle 1,0 \rangle * \begin{bmatrix} 3 \\ 0 \end{bmatrix} = 3 > 2$ True

2) if $u = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\langle 1,0 \rangle * \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2 = 2$ false

3) if $u = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\langle 1,0 \rangle * \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 < 2$ false

4) if $u = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$, $\langle 1,0 \rangle * \begin{bmatrix} 0 \\ 3 \end{bmatrix} = 0 < 2$ false

∴ u has to be $\begin{bmatrix} 3 \\ 0 \end{bmatrix}$

∴ all $x_1^i, x_2^i, x_3^i \in$ {the, a, this} and there would be 27 combination results for values of x_1^i, x_2^i, x_3^i.

Question 8:

using: $P(y=1 \mid x^i; W, b, V, \gamma) = \dfrac{\exp(V_1 \cdot h + \gamma_1)}{\exp(V_1 \cdot h + \gamma_1) + \exp(V_2 \cdot h + \gamma_2)}$

$P(y=2 \mid x^i; W, b, V, \gamma) = \dfrac{\exp(V_2 \cdot h + \gamma_2)}{\exp(V_1 \cdot h + \gamma_1) + \exp(V_2 \cdot h + \gamma_2)}$

If $x_1^i, x_2^i, x_3^i \in \{the, a, this\}$:

According to Question 7, we have $u = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$, $\therefore z = \langle 1, 0 \rangle \cdot \begin{bmatrix} 3 \\ 0 \end{bmatrix} - 2 = 1$

$\therefore h = g(1) = 1$

$\therefore \begin{cases} P(y=1 \mid x^i; W, b, V, \gamma) = 0.8 = \dfrac{\exp(V_1 + \gamma_1)}{\exp(V_1+\gamma_1) + \exp(V_2 + \gamma_2)} \\ P(y=2 \mid x^i; W, b, V, \gamma) = 0.2 = \dfrac{\exp(V_2 + \gamma_2)}{\exp(V_1 + \gamma_1) + \exp(V_2 + \gamma_2)} \end{cases} \Rightarrow \dfrac{\exp(V_1 + \gamma_1)}{\exp(V_2 + \gamma_2)} = 4 \quad ①$

If x_1^i, x_2^i and x_3^i at least one not from $\{the, a, this\}$:

According to Question 7, we have $u \neq \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ $\therefore W \cdot u + b \leq 0$

$\therefore h = 0$

$\therefore \begin{cases} P(y=1 \mid x^i; W, b, V, \gamma) = 0.5 = \dfrac{\exp(\gamma_1)}{\exp(\gamma_1) + \exp(\gamma_2)} \\ P(y=2 \mid x^i; W, b, V, \gamma) = 0.5 = \dfrac{\exp(\gamma_2)}{\exp(\gamma_1) + \exp(\gamma_2)} \end{cases} \Rightarrow \dfrac{\exp(\gamma_1)}{\exp(\gamma_1) + \exp(\gamma_2)} = \dfrac{\exp(\gamma_2)}{\exp(\gamma_1) + \exp(\gamma_2)}$

$\Rightarrow \exp(\gamma_1) = \exp(\gamma_2)$

$\Rightarrow \gamma_1 = \gamma_2 \quad ②$

$$\begin{cases} ① \\ ② \end{cases} \Rightarrow 4e^{V_2 + \gamma_2} = e^{V_1 + \gamma_1} = e^{V_1 + \gamma_2}$$

$$4 e^{V_2} \cancel{e^{\gamma_2}} = e^{V_1} \cancel{e^{\gamma_2}}$$

$$4 = e^{V_1 - V_2}$$

$$\ln 4 = V_1 - V_2$$

$\therefore V_1, V_2, \gamma_1, \gamma_2$ need to satisfy $\gamma_1 = \gamma_2$ and $V_1 - V_2 = \ln 4$.