CSCE636 Neural Network HW 4 Solution

1 Question 1

1.1 Question 1a

L should contain $|V| \times d$

H should contain $D \times D$ parameters.

I should contain $d \times D$ parameters.

b_1 should contain D parameters.

b_2 should contain $|V|$ parameters.

U should contain $D \times |V|$ parameters.

1.2 Question 1b

(1) $\frac{\partial E^{(t)}}{\partial U} = (h^{(t)})^T(y^{(t)} - \hat{y}^{(t)})$

(2) $\frac{\partial E^{(t)}}{\partial b_2} = y^{(t)} - \hat{y}^{(t)}$

(3) First, $\frac{\partial E^{(t)}}{\partial h^{(t)}} = (y^{(t)} - \hat{y}^{(t)})U^T$.
Let $a^{(t)} = h^{(t-1)}H + e^{(t)}I + b_1$,
then $\frac{\partial E^{(t)}}{\partial a^{(t)}} = \frac{\partial E^{(t)}}{\partial h^{(t)}} \odot sigmoid'(a^{(t)})$ where $\odot$ means element-wise multiplication.
We have $\frac{\partial E^{(t)}}{\partial I}|_{(t)} = (e^{(t)})^T \frac{\partial E^{(t)}}{\partial a^{(t)}}$

(4) $\frac{\partial E^{(t)}}{\partial H}|_{(t)} = (h^{(t-1)})^T \frac{\partial E^{(t)}}{\partial a^{(t)}}$

(5) $\frac{\partial E^{(t)}}{\partial b_1}|_{(t)} = \frac{\partial E^{(t)}}{\partial a^{(t)}}$

(6) $\frac{\partial E^{(t)}}{\partial h^{(t-1)}} = \frac{\partial E^{(t)}}{\partial a^{(t)}} H^T$

1.3 Question 1c

The cross-entropy and perplexity can be written as:

$PP^{(t)}(y^{(t)}\hat{y}^{(t)}) = \frac{1}{\hat{y}_k^{(t)}}$.

$CE^{(t)}(y^{(t)}\hat{y}^{(t)}) = -log(\hat{y}_k^{(t)}) = log(PP^{(t)}(y^{(t)}\hat{y}^{(t)})$.

2 Question 2

2.1 Question 2a

Single head: $d^2 + d^2 + d^2 = 3d^2$.

Multi head: $h \times 3 \times d^2/h = 3d^2$.

2.2 Question 2b

Single head: the total cost is $O(3nd^2 + n^2d + n^2d + n^2) = O(nd^2 + n^2d + n^2)$.

Multi head: the total cost is $O(h \times (3nd^2/h + n^2d/h + n^2d/h + n^2)) = O(nd^2 + n^2d + n^2h)$.

So, there is no significant difference between them.

3 Question 3

3.1 Question 3a

By assigning a self loop for each node. In other word, assigning 1 to each element on the diagonal of A.

$\hat{A} = A + I$

3.2 Question 3b

For each $a_{i,j}$ in A, it would be normalized as $a_{i,j} / \sum_{j=1}^{n} a_{i,j}$.