14. Word2Vec

1) Training Set, window size C = 2

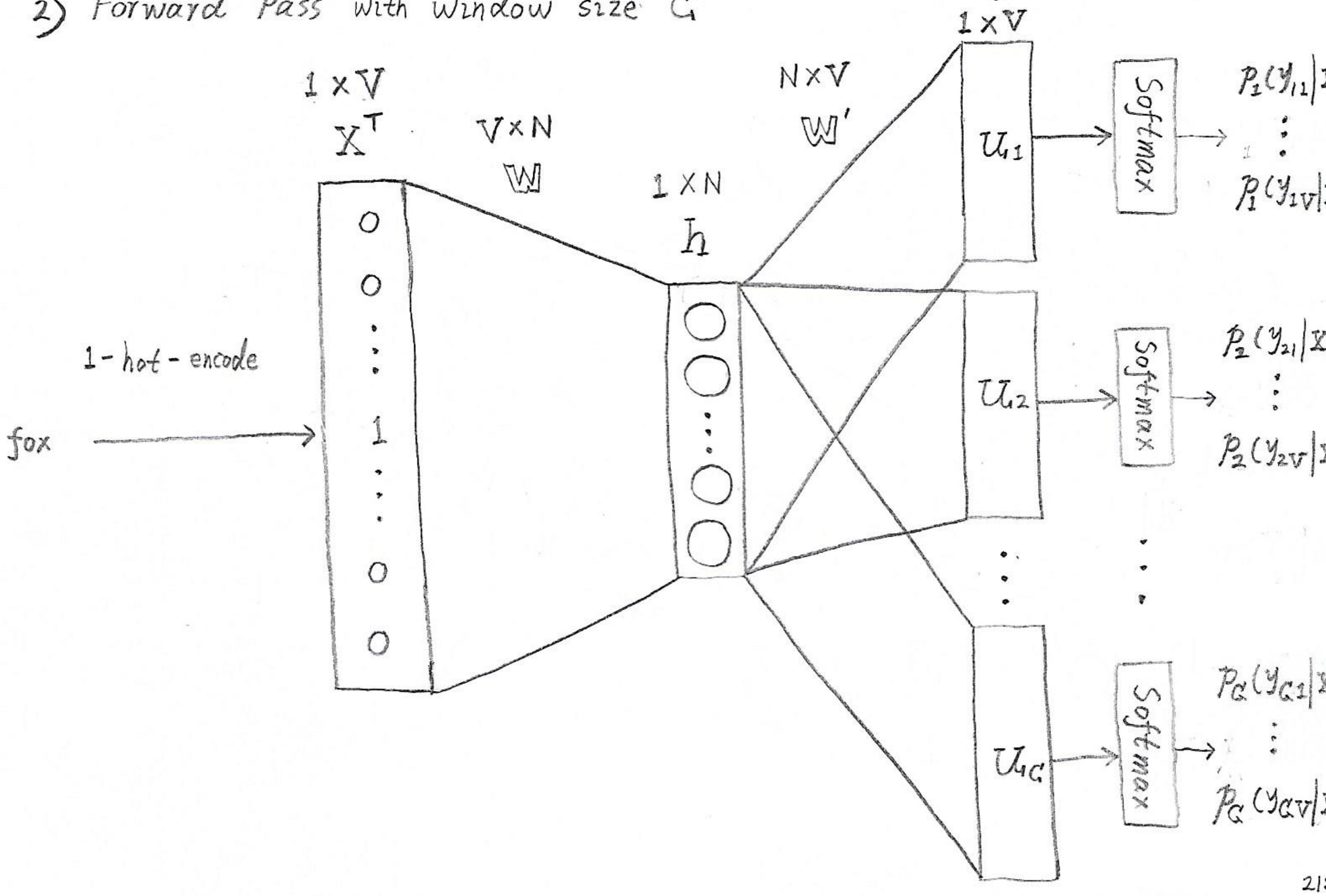
The quick brown fox jumps over the dog. → (The, quick), (The, brown)

The quick brown fox jumps over the dog. → (quick, The), (quick, brown), (quick, fox)

The quick brown fox jumps over the dog. → (brown, the), (brown, quick), (brown, fox), (brown, jumps)

The quick brown fox jumps over the dog. → (fox, quick), (fox, brown), (fox, jumps), (fox, over)

2) Forward Pass with window size C



$$\underset{1\times V}{X^T} \cdot \underset{V\times N}{\mathbb{W}}$$

$$= \begin{array}{c} \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \end{bmatrix} \\ \begin{array}{ccccccc} 1 & 2 & \cdots & k & \cdots & V-1 & V \end{array} \\ \uparrow \\ \text{fox} \end{array} \cdot \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & & \vdots \\ w_{k1} & w_{k2} & \cdots & w_{kN} \\ \vdots & \vdots & & \vdots \\ w_{V-1,1} & w_{V-1,2} & \cdots & w_{V-1,N} \\ w_{V1} & w_{V2} & \cdots & w_{VN} \end{bmatrix}$$

$$= \begin{bmatrix} w_{k1} & w_{k2} & \cdots & w_{kN} \end{bmatrix}$$

(Vector representation for "fox", aka, word2vec, to be learned by the neural network.)

$$= \underset{1\times N}{h}$$

$$\underset{1\times N}{h} \cdot \underset{N\times V}{\mathbb{W}'}$$

$$= \begin{bmatrix} h_1 & h_2 & \cdots & h_N \end{bmatrix} \cdot \begin{bmatrix} w'_{11} & w'_{12} & \cdots & \cdots & w'_{1V} \\ w'_{21} & w'_{22} & \cdots & \cdots & w'_{2V} \\ \vdots & \vdots & & & \vdots \\ w'_{N1} & w'_{N2} & \cdots & \cdots & w'_{NV} \end{bmatrix}$$

$$= \begin{bmatrix} h_1 \cdot w'_{11} + h_2 \cdot w'_{21} + \cdots + h_N \cdot w'_{N1} & h_1 \cdot w'_{12} + h_2 \cdot w'_{22} + \cdots + h_N \cdot w'_{N2} & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} u_1 & u_2 & \cdots & u_V \end{bmatrix}$$

$$= U_c, \quad c = 1, 2, \cdots, C$$

$$P_c = \text{Softmax}(X^T \cdot \mathbb{W} \cdot \mathbb{W}')$$

$$= \text{Softmax}(U_c)$$

$$= \left[\text{Softmax}(u_{c1}) \quad \text{Softmax}(u_{c2}) \quad \cdots \quad \text{Softmax}(u_{cV}) \right]$$

$$= \left[\frac{e^{u_{c1}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \quad \frac{e^{u_{c2}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \quad \cdots \quad \frac{e^{u_{cV}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \right]$$

$$= \left[p_c(y_{c1} | X) \quad p_c(y_{c2} | X) \quad \cdots \quad p_c(y_{cV} | X) \right], \text{ where}$$

$$Y_c = \begin{bmatrix} y_{c1} \\ \vdots \\ y_{cj} \\ \vdots \\ y_{cV} \end{bmatrix} \text{Estimated Outputs at C}, \quad \underbrace{Y'_c}_{\text{True outputs at C}} = \begin{bmatrix} y'_{c1} \\ \vdots \\ y'_{cj} \\ \vdots \\ y'_{cV} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \xrightarrow{} \text{brown}, \quad c = 1, 2, \cdots, C$$

$$L = \sum_{c=1}^{C} L_c$$

Cth True output at index i. Cth Estimated output at index i

$$= \sum_{c=1}^{C} \left(- \sum_{i=1}^{V} y'_{ci} \cdot \log(p_c(y_{ci} | X)) \right) \quad \text{(Cross Entropy Loss)}$$

$$= \sum_{c=1}^{C} \left(- y'_{cj} \cdot \log(p_c(y_{cj} | X)) \right) \quad (\text{If } y'_{ci} \neq y'_{cj}, \text{ then } y'_{ci} = 0$$

$$= \sum_{c=1}^{C} \left(- \log(p_c(y_{cj} | X)) \right) \quad (y'_{cj} = 1)$$

$$= - \sum_{c=1}^{C} \log\left(\frac{e^{u_{cj}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \right)$$

$$= -\left(\sum_{C=1}^{C} \left(\log e^{u_{cj}} - \log \sum_{j'=1}^{V} e^{u_{cj'}} \right) \right)$$

$$= - \sum_{C=1}^{C} u_{cj} + \sum_{C=1}^{C} \cdot \log \sum_{j'=1}^{V} e^{u_{cj'}},$$

3) Backward Pass with window size C

$$\frac{\partial L}{\partial u_{ci}} = \frac{\partial L}{\partial u_{ci}} \left(- \sum_{C=1}^{C} u_{cj} + \sum_{C=1}^{C} \log \sum_{j'=1}^{V} e^{u_{cj'}} \right)$$

$$= - \frac{\partial L}{\partial u_{ci}} u_{cj} + \frac{\partial L}{\partial u_{ci}} \log \sum_{j'=1}^{V} e^{u_{cj'}}$$

$$= - \frac{\partial L}{\partial u_{ci}} u_{cj} + \frac{1}{\sum_{j'=1}^{V} e^{u_{cj'}}} \cdot \frac{\partial L}{\partial u_{ci}} \left(\sum_{j'=1}^{V} e^{u_{cj'}} \right)$$

When $u_{ci} = u_{cj}$,

$$\frac{\partial L}{\partial u_{cj}} = -1 + \frac{e^{u_{cj}}}{\sum_{j'=1}^{V} e^{u_{cj'}}}$$

When $u_{ci} \neq u_{cj}$

$$\frac{\partial L}{\partial u_{ci}} = \frac{e^{u_{ci}}}{\sum_{j'=1}^{V} e^{u_{cj'}}}$$

$$\frac{\partial L}{\partial u_{ci}} = \begin{bmatrix} \dfrac{e^{u_{c1}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \\ \vdots \\ \dfrac{e^{u_{cj}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \\ \vdots \\ \dfrac{e^{u_{cV}}}{\sum_{j'=1}^{V} e^{u_{cj'}}} \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} p_c(y_{c1} | X) \\ \vdots \\ p_c(y_{cj} | X) \\ \vdots \\ p_c(y_{cV} | X) \end{bmatrix} - \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$= p_c(y_{ci} | X) - 1_{u_{ci} = u_{cj}}$$

$$\frac{\partial L}{\partial \mathbb{W}'} = \frac{\partial L}{\partial \mathbb{U}_1} \cdot \frac{\partial \mathbb{U}_1}{\partial \mathbb{W}'} + \frac{\partial L}{\partial \mathbb{U}_2} \cdot \frac{\partial \mathbb{U}_2}{\partial \mathbb{W}'} + \cdots + \frac{\partial L}{\partial \mathbb{U}_C} \cdot \frac{\partial \mathbb{U}_C}{\partial \mathbb{W}'}$$

$$= \sum_{c=1}^{C} \frac{\partial L}{\partial \mathbb{U}_c} \cdot \frac{\partial \mathbb{U}_c}{\partial \mathbb{W}'}$$

$$\frac{\partial L}{\partial w'_{ni}} = \sum_{c=1}^{C} \frac{\partial L}{\partial u_{ci}} \cdot \frac{\partial u_{ci}}{\partial w'_{ni}} = \sum_{c=1}^{C} \frac{\partial L}{\partial u_{ci}} \cdot \frac{\partial}{\partial w'_{ni}} \left(\sum_{n=1}^{N} h_n \cdot w'_{ni} \right)$$

$$= \sum_{c=1}^{C} \left(p_c(y_{ci} | X) - 1_{u_{ci} = u_{cj}} \right) \cdot h_n$$

$$= \sum_{c=1}^{C} \left(p_c(y_{ci} | X) - 1_{u_{ci} = u_{cj}} \right) \cdot \left(\sum_{m=1}^{V} x_m \cdot w_{mn} \right)$$

where $n = 1, 2, \cdots, N, \quad i = 1, 2, \cdots, V$

$$\frac{\partial L}{\partial h_n} = \left(\frac{\partial L}{\partial u_{11}} \cdot \frac{\partial u_{11}}{\partial h_n} + \frac{\partial L}{\partial u_{12}} \cdot \frac{\partial u_{12}}{\partial h_n} + \cdots + \frac{\partial L}{\partial u_{1V}} \cdot \frac{\partial u_{1V}}{\partial h_n}\right) +$$

$$\vdots$$

$$\left(\frac{\partial L}{\partial u_{C1}} \cdot \frac{\partial u_{C1}}{\partial h_n} + \frac{\partial L}{\partial u_{C2}} \cdot \frac{\partial u_{C2}}{\partial h_n} + \cdots + \frac{\partial L}{\partial u_{CV}} \cdot \frac{\partial u_{CV}}{\partial h}\right)$$

$$= \sum_{i=1}^{V} \frac{\partial L}{\partial u_{1i}} \cdot \frac{\partial u_{1i}}{\partial h_n} + \cdots + \sum_{i=1}^{V} \frac{\partial L}{\partial u_{Ci}} \cdot \frac{\partial u_{Ci}}{\partial h_n}$$

$$= \sum_{C=1}^{C} \sum_{i=1}^{V} \frac{\partial L}{\partial u_{ci}} \cdot \frac{\partial u_{ci}}{\partial h_n} = \sum_{C=1}^{C} \sum_{i=1}^{V} \frac{\partial L}{\partial u_{ci}} \cdot \frac{\partial}{\partial h_n}\left(\sum_{n=1}^{N} h_n \cdot w'_{ni}\right)$$

$$= \sum_{C=1}^{C} \sum_{i=1}^{V} \left(P_c(y_{ci} \mid X) - 1_{u_{ci} = u_{cj}}\right) \cdot w'_{ni}, \text{ where}$$

$n = 1, 2, \cdots, N$

$$\frac{\partial L}{\partial w_{mn}} = \frac{\partial L}{\partial h_n} \cdot \frac{\partial h_n}{\partial w_{mn}}$$

$$= \frac{\partial L}{\partial h_n} \cdot \frac{\partial}{\partial w_{mn}}\left(\sum_{m=1}^{V} x_m \cdot w_{mn}\right)$$

$$= \frac{\partial L}{\partial h_n} \cdot x_m$$

$$= \sum_{C=1}^{C} \sum_{i=1}^{V} \left(P_c(y_{ci} \mid X) - 1_{u_{ci} = u_{cj}}\right) \cdot w'_{ni} \cdot x_m, \text{ where}$$

$m = 1, 2, \cdots, V$, and this is used to update Word2Vec $\mathbb{W}$.