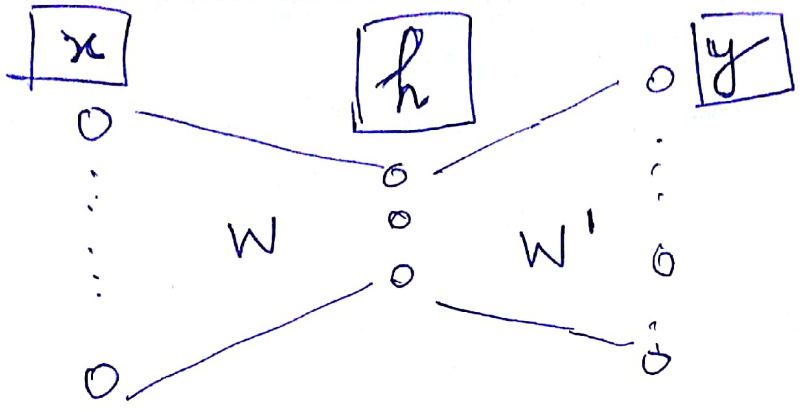
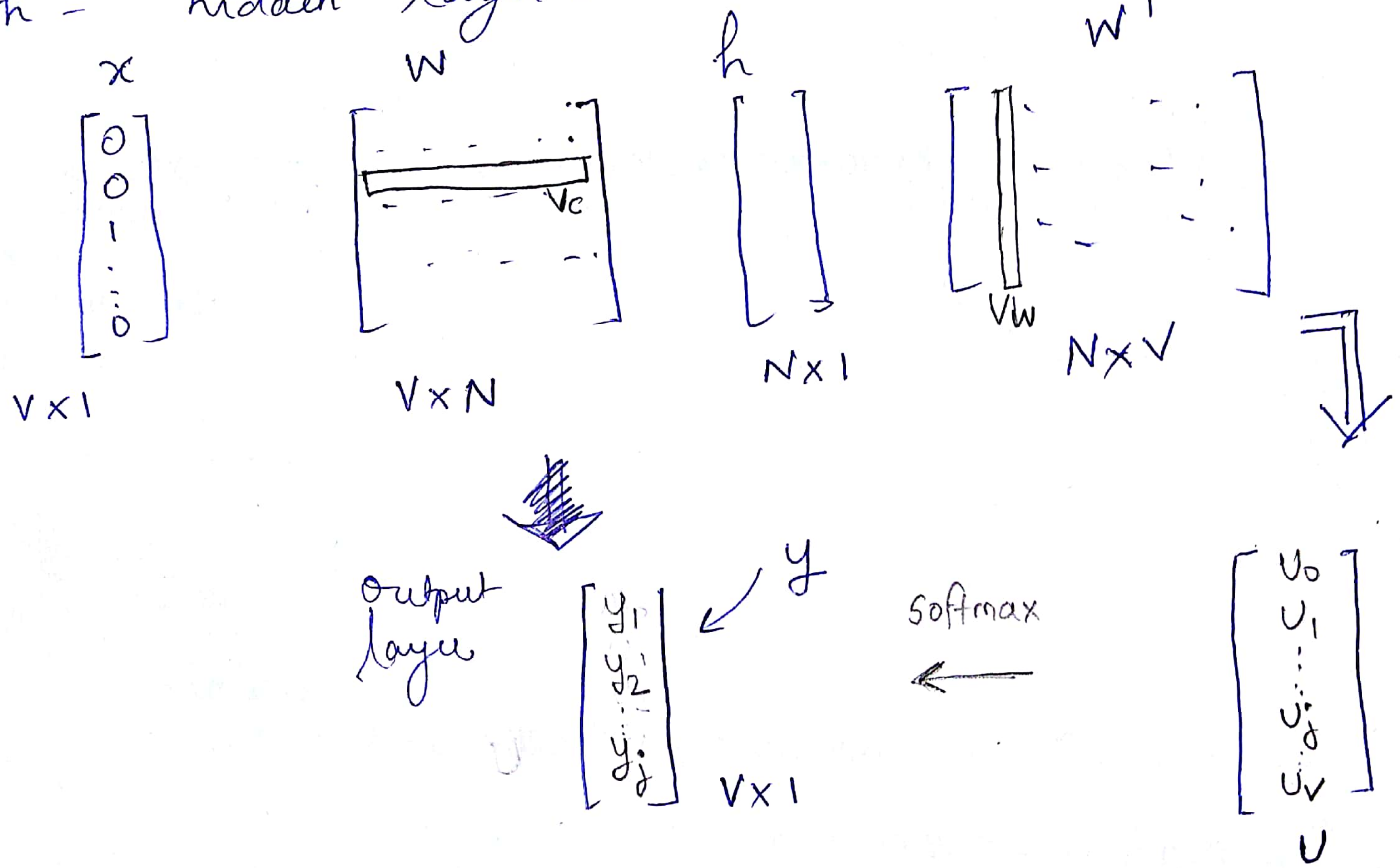
Word Embeddings - Skip Gram model.



x - one hot representation of word w (the context word)

h - hidden layer.



$h = W^T x$ $\{x^T W$ is a row of matrix $W\}$

so h is a row of W (a row of context word)

$\boxed{h = v_c = W^T x}$ $\{\because W^T x$ is of $N \times 1$ = size of $h\}$

$$\underset{(V\times 1)}{U} = \underset{(V\times N)}{W'^T}\ \underset{(N\times 1)}{h}$$ → matrix representation

$$\cancel{u_j = V_{w_j}^T \cdot V_{c_j}}$$ → scalar representation

$$u_j = \underset{1\times N}{V_{w_j}^T} \cdot \underset{N\times 1}{V_c^T},\ h = V_{w_j}^T h$$

$$y = \text{softmax}(U)$$

$$y_j = \frac{e^{u_j}}{\sum_{j'=1}^{V} e^{u_{j'}}}$$

$$P(w_j \mid w_i) = y_j = \frac{e^{u_j}}{\sum_{j'=1}^{V} e^{u_{j'}}} = \frac{e^{V_{w_j}^T \cdot h}}{\sum_{j'=1}^{V} e^{V_{w_{j'}}^T h}}$$

Goal maximise y_{j^*}

j^* = index of content words.

$$E = -\log y_{j^*} = -\log\left(\frac{e^{u_{j^*}}}{\sum_{j'=1}^{V} e^{u_{j'}}}\right)$$

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

$$E = -u_{j^*} + \log\sum_{j'=1}^{V} e^{u_{j'}}$$

$$E = -\log\left\{\prod_{c=1}^{C} \frac{e^{u_{j_c^*} \cdot t_{j_c^*}}}{\sum_{j'=1}^{V} e^{u_{j'}}}\right\}$$

$$E = -\log\left\{\prod_{c=1}^{C} e^{u_{j_c^*} \cdot t_{j^*c}}\right\} + \left(\log\sum_{j'=1}^{V} e^{u_{j'}}\right)^{C}$$

$$E = -\sum_{c=1}^{C} u_{j_c^*} \cdot t_{j^*c} + C\log\sum_{j'=1}^{V} e^{u_{j'}}$$

$$\boxed{E = -\sum_{c=1}^{C} u_{j_c^*} \cdot t_{j^*c} + C\log\sum_{j'=1}^{V} e^{u_{j'}}}$$

$$\frac{\partial E}{\partial u_j} \times \frac{\partial u_j}{\partial w'_{ij}} = \frac{\partial E}{\partial w'_{ij}}$$

$$\frac{\partial E}{\partial u_j} = -\sum_{c=1}^{C} t_{j^*c} + C \frac{1}{\sum_{j'=1}^{V} e^{u_{j'}}} \frac{\partial}{\partial u_j} \sum_{j=1}^{V} e^{u_j}$$

$$= -\sum_{c=1}^{C} t_{j^*_c} + \sum y_j = y_j - t_j = e_j$$

$$\frac{\partial E}{\partial w'_{ij}} = e_j \times \frac{\partial u_j}{\partial w'_{ij}} = e_j \times \frac{\partial (w'_{ij} \times h_i)}{\partial w'_{ij}} = e_j \times h_i$$

$\cancel{u_j = w'_{ij} \times h_i}$

$$\frac{\partial E}{\partial w'_{ij}} = e_j \times h_i$$

$\cancel{(V \times 1)\ (1 \times N)}$

W'	$h_i \quad e_j^T$
$(N \times V)$	$N \times 1 \quad 1 \times V$

$h_i = W_j^T x_k$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial u_j} \frac{\partial u_j}{\partial w_{ij}} \cancel{= e_j \times \partial}$$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial u_j} \times \frac{\partial u_j}{\partial h_i} \times \frac{\partial h_i}{\partial w_{ij}}$$

$$= e_j \times w'_{ij} \times \frac{\partial (w_{ij} \times x_k)}{\partial w_{ij}} = e_j \times w'_{ij} \times x_k$$

$\{x_k = 1\}$

$$\frac{\partial E}{\partial w_{ij}} = e_j \times w'_{ij} \times x_k$$

e	w'	x
V×1	N×V	V×1

$w'^T \times e$

$\cancel{V \times N}$

$N \times V$

$$\underbrace{\underset{N \times V}{w'} \quad \underset{V \times 1}{e}}_{N \times 1} \quad \underbrace{\underset{\cancel{V \times 1}\ 1 \times V}{x^T}}_{1 \times V}$$

$$\underbrace{N \times 1 \quad 1 \times V}_{N \times V}$$

$\underline{N \times V} \Leftarrow$ dimensions of W