Assignment 1.

1.(a). $softmax(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$

$$\Rightarrow softmax(x_i + c)$$
$$= \frac{e^{x_i+c}}{\sum_j e^{x_j+c}} = \frac{e^{x_i} \cdot e^c}{\sum_j e^{x_j} \cdot e^c} = \frac{e^c \cdot e^{x_i}}{e^c \sum_j e^{x_j}}$$
$$= \frac{e^{x_i}}{\sum_j e^{x_j}} = softmax(x_i)$$
$$\Rightarrow softmax(x) = softmax(x+c).$$

2.(a). $\sigma(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$

$$\Rightarrow \sigma'(x) = -(1+e^{-x})^{-2} \cdot (-e^{-x})$$
$$= \frac{e^{-x}}{(1+e^{-x})^2}$$
$$= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}}$$
$$= \frac{1}{1+e^{-x}} \cdot \left(1 - \frac{1}{1+e^{-x}}\right)$$
$$= \sigma(x)(1-\sigma(x))$$

(b) 已知 y 为 one-hot vector. 令 $y_k = 1$. $y_{i \neq k} = 0$.

则 $CE(y, \hat{y}) = -\sum_i y_i \log(\hat{y}_i)$
$$= -y_k \log(\hat{y}_k)$$
$$= -\log(\hat{y}_k).$$

$$\Rightarrow \frac{\partial CE(y,\hat{y})}{\partial \theta_i} = \frac{\partial(-\log(\hat{y}_k))}{\partial \theta_i} = \frac{\partial\left(-\log \frac{e^{\theta_k}}{\sum_j e^{\theta_j}}\right)}{\partial \theta_i}$$
$$= \frac{\partial\left(\log \sum_j e^{\theta_j} - \log e^{\theta_k}\right)}{\partial \theta_i}$$

$$= \frac{\partial \log \sum_j e^{\theta_j}}{\partial \theta_i} - \frac{\partial \theta_k}{\partial \theta_i}$$

$$= \begin{cases} \hat{y}_i & i \neq k \\ \hat{y}_i - 1 & i = k. \end{cases}$$

$$\Rightarrow \frac{\partial CE(y, \hat{y})}{\partial \theta} = \hat{y} - y$$

(c) 令 one-hot vecter y 中 $y_k = 1$ $y_{i\neq k} = 0$.

则 $J = CE(y, \hat{y}) = -\log(\hat{y}_k)$

令 $\theta_1 = xW_1 + b_1$

$\theta_2 = hW_2 + b_1$

$$\Rightarrow \frac{\partial J}{\partial \theta_2} = \hat{y} - y. \quad \Rightarrow \frac{\partial J}{\partial h} = \frac{\partial J}{\partial \theta_2}\frac{\partial \theta_2}{\partial h} = (\hat{y} - y)W_2^T$$

$$\Rightarrow \frac{\partial J}{\partial \theta_1} = \frac{\partial J}{\partial h}\frac{\partial h}{\partial \theta_1} = (\hat{y} - y)W_2^T \sigma'(\theta_1)$$

$$\Rightarrow \frac{\partial J}{\partial x} = \frac{\partial J}{\partial \theta_1}\frac{\partial \theta_1}{\partial x} = (\hat{y} - y)W_2^T \times \sigma(\theta_1)W_1^T$$

(d) $W_1: D_x * H.$ $\quad b_1: 1 * H$

$W_2: H * D_y.$ $\quad b_2: 1 * D_y$

Total: $\|\theta\|$: $D_x H + H + D_y H + D_y$

$= (D_x + 1)H + (H + 1)D_y.$

3.(a) 对于 $J_{softmax-CE}(o, v_c, U) = CE(y, \hat{y})$

有 $\theta = U^T v_c$　　$(U: n * |D|$　$v_c: n * 1$

$$\Rightarrow \frac{\partial J}{\partial v_c} = \frac{\partial J}{\partial \theta} * \frac{\partial \theta}{\partial v_c} = U^T(\hat{y} - y)$$

(b). 令 $\theta = U^T v_c$.

$$\frac{\partial J}{\partial U} \Rightarrow \frac{\partial J}{\partial u_w} = \frac{\partial J}{\partial \theta_k} \cdot \frac{\partial \theta_k}{\partial u_w} = (\hat{y} - y) \frac{\partial (u_k^T \cdot v_c)}{\partial u_w}$$

$$= (\hat{y} - y) \cdot v_c \longrightarrow \begin{cases} (\hat{y} - 1) v_c & k = w \\ \hat{y} v_c & k \neq w. \end{cases}$$

(c) 对于 $J_{neg-sample}(o, v_c, U) = -\log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K} \log(\sigma(-u_k^T v_c))$

由 sigmoid 性质 $\log \sigma = 1 - \sigma$, $\sigma' = \sigma(1-\sigma)$

$$\frac{\partial J}{\partial v_c} = (\sigma(u_o^T v_c) - 1) u_o - \sum_{k=1}^{K} (\sigma(-u_k^T v_c) - 1) u_k.$$

$$\frac{\partial J}{\partial u_w} = -\frac{\partial \log \sigma(u_o^T v_c)}{\partial u_w} - \frac{\sum_k \partial \log \sigma(-u_k^T v_c)}{\partial u_w}$$

$$= \begin{cases} (\sigma(u_o^T v_c) - 1) v_c & w = o \\ -(\sigma(-u_k^T v_c) - 1) v_c. & w = k \\ 0 & \text{else} \end{cases}$$

其中 K 为负采样样本.

(d) 对于 skip-gram:

$$\frac{\partial J_{skip\text{-}gram}}{\partial v_k} = \sum_{j\in[-m,m],\, j\neq 0} \frac{\partial F(w_{c+j}, v_c)}{\partial v_k}$$

$$\frac{\partial J_{skipgram}}{\partial u_k} = \sum_{j\in[-m,m],\, j\neq 0} \frac{\partial F(w_{c+j}, v_c)}{\partial u_k}$$

对于 CBOW

$$\frac{\partial J_{CBOW}}{\partial v_k} = \frac{\partial F(w_c, \hat{v})}{\partial v_k} = \frac{\partial F(w_c, \hat{v})}{\partial \hat{v}} \frac{\partial \hat{v}}{\partial v_k}$$

$$= \begin{cases} \dfrac{\partial F(w_c, \hat{v})}{\partial \hat{v}} & k \in \{w_{c+j},\ j\in[-m,m],\ j\neq 0\} \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial J_{CBOW}}{\partial u_k} = \frac{\partial F(w_c, \hat{v})}{\partial u_k}$$

$$\Rightarrow \frac{\partial J_{CBOW}}{\partial U} = \frac{\partial F(w_c, \hat{v})}{\partial U}$$