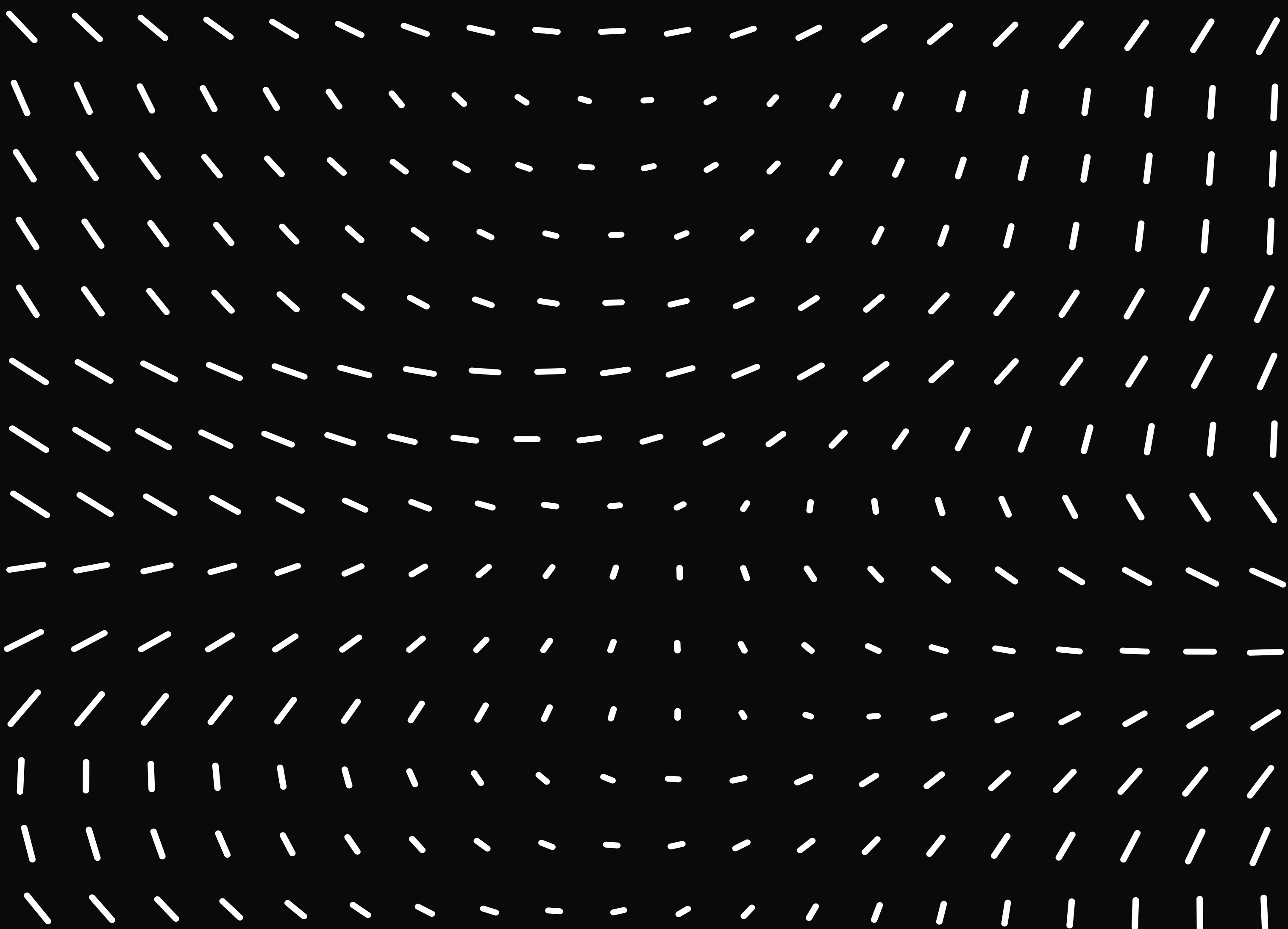
Derivations for the Assignment - 2

Name:- D. Varun Kumar Reddy

CWID:- 10474196



① we know that $\sigma(x) = \frac{1}{1+e^{-x}}$

so $\frac{\partial}{\partial x}\sigma(x) = \frac{\partial}{\partial x}\left[\frac{1}{1+e^{-x}}\right]$

$\Rightarrow \frac{\partial}{\partial x}\left[(1+e^{-x})^{-1}\right]$

$\Rightarrow \frac{\partial}{\partial(1+e^{-x})}\left[(1+e^{-x})^{-1}\right] \cdot \frac{\partial}{\partial x}(1+e^{-x})$

$\Rightarrow (1+e^{-x})^{-2} \times (-1) \times (-e^{-x})$

$\Rightarrow \frac{e^{-x}}{(1+e^{-x})^2}$

$\Rightarrow \left(\frac{1}{1+e^{-x}}\right)\left(\frac{e^{-x}}{1+e^{-x}}\right)$

$\Rightarrow \left(\frac{1}{1+e^{-x}}\right)\left(1 - \frac{1}{1+e^{-x}}\right)$

$\Rightarrow \sigma(x)\,(1-\sigma(x))$

so $\frac{\partial}{\partial x}\sigma(x) = \sigma(x)\,(1-\sigma(x))$

② $\mathcal{L} = -\sum_i y_i \log(\hat{y}_i)$

$= -\sum_i y_i \log(p(o/c))$

now $\frac{\partial \mathcal{L}}{\partial v_c} = \frac{\partial}{\partial v_c}\left[-\log\left(\frac{e^{u_o^T v_c}}{\sum^W e^{u_w^T v_c}}\right)\right]$

$= -\frac{\partial}{\partial v_c}\left[u_o^T v_c - \log\left(\sum_1^W e^{u_w^T v_c}\right)\right]$

$= -\left[u_o^T - \frac{1}{\sum_{w=1}^W e^{u_w^T v_c}} \cdot \frac{\partial}{\partial v_c}\sum e^{u_w^T v_c}\right]$

$= -\left[u_o^T - \frac{1}{\sum_{w=1}^W \exp(u_w^T v_c)} \cdot \sum \exp(u_w^T v_c) \cdot u_w^T\right]$

$= -\left[u_o^T - \frac{\sum_{o=1}^W \exp(u_w^T v_c)}{\sum_{w=1}^W \exp(u_w^T v_c)} \cdot u_w^T\right]$

$= -\left[u_o^T - \left(\sum_{o=1}^W \hat{y}_o\right) \cdot u_w^T\right]$

$= -\left[u_o^T - \sum_{w=1}^W u_w^T \hat{y}_w\right]$

$= \sum_{w=1}^W u_w^T \hat{y}_w - u_o^T$

$$= \quad U \cdot \hat{y} - U \cdot y$$

$$= \quad u(\hat{y} - y)$$

so gradient descent of the $\sum y_i \log(p(o/c))$ is $\boxed{u(\hat{y}-y)}$

③ now $\frac{\partial L}{\partial v_w} = \frac{\partial L}{\partial v} = \begin{cases} Ⓐ, & w = o \\ Ⓑ, & o/w \end{cases} = \begin{cases} v_c[\hat{y}_o - 1], & w = o \\ \hat{y}_w, & o/w \end{cases}$

Ⓐ $\frac{\partial L}{\partial u_o} = -\frac{\partial}{\partial u_o}\left[u_o^T v_c - \log\left(\sum \exp(u_w^T \cdot v_c)\right)\right]$

$$= -\left[v_c - \frac{1}{\sum \exp(v_w^T \cdot v_c)} \cdot \frac{\partial}{\partial u_o} \sum \exp(u_w^T \cdot v_c)\right]$$

$$= -\left[v_c - \frac{1}{\sum \exp(u_w^T v_c)} \cdot \sum \exp(u_w^T \cdot v_c) \cdot v_c\right]$$

$$= -\left[v_c - v_c \hat{y}_o\right]$$

$$= \boxed{v_c[\hat{y}_o - 1]}$$

Ⓑ $\frac{\partial L}{\partial u_w} = -\frac{\partial}{\partial u_w}\left[u_o^T v_c - \log\left(\sum \exp(u_w^T \cdot v_c)\right)\right]$

$$= -\left[0 - \frac{1}{\sum \exp(v_w^T \cdot v_c)} \cdot \sum \exp(v_w^T \cdot w) \cdot v_c\right]$$

$$= -\left[0 - \hat{y}_w\right]$$

$$= \boxed{\hat{y}_w.}$$

(4)

$$J_{\text{neg-sample}}(o, \mathbf{v}_c, U) = -\log(\sigma(\mathbf{u}_o^\top \mathbf{v}_c)) - \sum_{k=1}^{K} \log(\sigma(-\mathbf{u}_k^\top \mathbf{v}_c))$$

now

$$\frac{\partial J_{\text{neg-sample}}}{\partial v_c} = \frac{\partial}{\partial v_c}\left[-\log(\sigma(u_o^T v_c)) - \sum_{k=1}^{K} \log(\sigma(-u_k^T v_c)) \right]$$

$$= -\left[\frac{1}{\cancel{\sigma(u_o^T v_c)}} \cdot \cancel{\sigma(u_o^T v_c)}. (1-\sigma(u_o^T v_c)). u_o^T + \sum_{k=1}^{K} \frac{1}{\cancel{\sigma(-u_k^T v_c)}} \cdot \cancel{\sigma(-u_k^T v_c)} (1-\sigma(-u_k^T \cdot v_c)) u_k^T \right]$$

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

$$\begin{cases} (A), & w=0 \\ (B), & o/w \end{cases} = \begin{cases} -(1-\sigma(u_o^T v_c))\, v_c, & w=0 \\ (1-\sigma(-u_w^T v_c)) v_c, & o/w \end{cases}$$

(5) Given $J = \sum_{\substack{-m \le j \le m \\ j \neq 0}} F(w_{c+j}, v_c)$

where $F(w_{c+j}, v_c)$ can be $J_{smx\text{-}CE}$ (or) $J_{neg\text{-}samp\text{-}CE}$.

then, $$\frac{\partial J}{\partial \xi} = \begin{cases} (a), & \xi \in U = [u_1, u_2, \dots\dots, u_w] \\ (b), & \xi = v_c \\ (c), & o/w \end{cases}$$

$$(a) = \frac{\partial}{\partial U} \Sigma F = \sum_{\substack{-m \le j \le m \\ j \neq 0}} \frac{\partial F(w_{c+j}, v_c)}{\partial U} = \frac{\partial L}{\partial U_w}$$

(smx (or) neg-sample) ↳ from part (c) 4 (d)

$$(b) = \frac{\partial}{\partial v_c} \Sigma F = \sum_{\substack{-m \le j \le m \\ j \neq 0}} \frac{\partial F(w_{c+j}, v_c)}{\partial v_c} = \frac{\partial L}{\partial v_c}$$

(smx (or) neg-sample) ↳ from part (a) 4 (b)

$$\frac{\partial J}{\partial \xi} = \begin{cases} v_c[\hat{y}_o - 1] & , w=0, \text{smx CE}, \xi \in U \\ \hat{y}_w & , o/w, \text{smx CE}, \xi \in U \\ -(1-\sigma(u_o^T v_c)) v_c & , w=0, \text{neg-CE}, \xi \in U \\ (1-\sigma(-u_o^T v_c)) v_c & , o/w, \text{neg-CE}, \xi \in U \\ \sum_{w=1}^{W} u_w^T \hat{y}_w - u_o^T & , \text{smx CE}, \xi = v_c \\ \sum_{1}^{K} u_k^T (1-\sigma(-u_k^T v_c)) - (1-\sigma(u_o^T v_c)).u_o^T & , \text{neg-CE}, \xi = v_c \\ 0 & , o/w \end{cases}$$