Extra Credit Challenge-I

$$J_{naive\text{-}softmax}(v_c, o, U) = -u_o^T v_c + \log\left(\sum_{w \in Vocab} \exp(u_w^T v_c)\right)$$

Now, we know, that,

$$\frac{d}{dx}\log(f(x)) = \frac{f'(x)}{f(x)} \quad \text{and,} \quad \frac{d}{dx}\exp(f(x)) = \exp(f(x)) f'(x)$$

$$\text{so,} \quad \frac{d}{dv_c}\log\left(\sum_{w=1}^{V} \exp(u_w^T v_c)\right) = \frac{\sum_{w=1}^{V} \exp(u_w^T v_c) u_w^T}{\sum_{x=1}^{V} \exp(u_x^T v_c)}$$

$$\text{so,} \quad \frac{dJ}{dv_c} = -u_o^T + \frac{\sum_{w=1}^{V} \exp(u_w^T v_c)\, u_w^T}{\sum_{x=1}^{V} \exp(u_x^T v_c)}$$

$$= -u_o^T + \sum_{w=1}^{V} \frac{\exp(u_w^T v_c)}{\sum_{x=1}^{V} \exp(u_x^T v_c)} u_w^T$$

$$= -u_o^T + \sum_{w \in Vocab} \hat{y}_w u_w^T$$

Now, y is one hot vector, with, $y_o = 1$, so, in ~~matrix~~ notation

$$-u_o^T = U.y$$

$$\text{and,} \quad \sum_{w \in Vocab} \hat{y}_w u_w^T = U.\hat{y}$$

$$\text{so,} \quad \frac{dJ}{dv_c} = U(\hat{y} - y)$$

Extra Credit Challenge-II

$$J_{naive\_softmax}(v_c, o, U) = -u_o^T v_c + \log\left(\sum_{w' \in vocab} \exp(u_{w'}^T v_c)\right)$$

Now, we know, that,

$$\frac{d}{dx}(\log(f(x))) = \frac{f'(x)}{f(x)} \text{ and, } \frac{d}{dx}(\exp(f(x))) = \exp(f(x))\,f'(x)$$

$$\text{and, } \frac{d}{dx}(\text{constant}) = 0$$

so, when, $w = o$,

$$\frac{d}{du_w}\left(\log\left(\sum_{w' \in vocab} \exp(u_{w'}^T v_c)\right)\right) = \frac{\exp(u_w^T v_c)\, v_c}{\sum_{x=1}^{V} \exp(u_x^T v_c)} = \hat{y}_w v_c$$

$$\text{so, } \frac{dJ}{du_w} = -v_c + \hat{y}_w v_c$$

Now, y is one hot vector with $y_o = y_w = 1$, so, in matrix notation,

$$\frac{dJ}{du_w} = (\hat{y}_w - 1)\, v_c \quad \text{if } w = o \;=\; v_c(\hat{y} - y)^T$$

when, $w \neq o$,

$$\frac{dJ}{du_w} = \hat{y}_w v_c \quad \text{as, } \frac{d}{du_w}(-u_o^T v_c) = 0$$

so, this, can also be written as,

$$\frac{dJ}{du_w} = v_c(\hat{y} - y)^T.$$