Given:

$$J_{naive-softmax}(v_c, o, U) = -\sum_{w \in Vocab} y_w \log(\hat{y}_w) = -\log(\hat{y}_o) = -\log\left(\frac{\exp({u_o}^T v_c)}{\sum_{w \in Vocab} \exp({u_w}^T v_c)}\right)$$

$$= -\left(\log(\exp({u_o}^T v_c)) - \log\left(\sum_{w \in Vocab} \exp({u_w}^T v_c)\right)\right)$$

$$= -{u_o}^T v_c + \log\left(\sum_{w \in Vocab} \exp({u_w}^T v_c)\right)$$

Extra Credit Challenge I (2.5 Points)

Write the Steps to arrive at $\frac{\partial J}{\partial v_c} = -u_o + \sum_{w=1}^{V} \hat{y}_w u_w$

$$\frac{\partial J}{\partial v_c} = \frac{\partial}{\partial v_c}(-{u_o}^T v_c) + \frac{\partial}{\partial v_c}\log\left(\sum_{w \in Vocab} \exp({u_w}^T v_c)\right)$$

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

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

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

$$= -u_o + \sum_{x \in Vocab} \frac{\exp({u_x}^T v_c)}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} u_x = -u_o + \sum_{x \in Vocab} \hat{y}_x u_x$$

Extra Credit Challenge II (2.5 Points)

Write the Steps to arrive at $\frac{\partial J}{\partial u_w} = \begin{cases} (\hat{y}_w - 1)v_c & if\ w = o \\ \hat{y}_w v_c & otherwise \end{cases}$

If $w = o$

$$\frac{\partial J}{\partial u_w} = \frac{\partial J}{\partial u_o} = \frac{\partial}{\partial u_o}(-{u_o}^T v_c) + \frac{\partial}{\partial u_o}\log\left(\sum_{w \in Vocab} \exp({u_w}^T v_c)\right)$$
$$= -v_c + \frac{1}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} \sum_{x \in Vocab} \frac{\partial}{\partial u_o} \exp({u_x}^T v_c)$$
$$= -v_c + \frac{1}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} \sum_{x \in Vocab} \exp({u_x}^T v_c) \frac{\partial}{\partial u_o} {u_x}^T v_c$$
$$= -v_c + \frac{\exp({u_o}^T v_c)}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} v_c = -v_c + \hat{y}_o v_c = (\hat{y}_o - 1)v_c = (\hat{y}_w - 1)v_c$$

If $w \neq o$

$$\frac{\partial J}{\partial u_w} = \frac{\partial}{\partial u_w}(-{u_o}^T v_c) + \frac{\partial}{\partial u_w}\log\left(\sum_{w \in Vocab} \exp({u_w}^T v_c)\right)$$
$$= \frac{1}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} \sum_{x \in Vocab} \frac{\partial}{\partial u_w} \exp({u_x}^T v_c)$$
$$= \frac{1}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} \sum_{x \in Vocab} \exp({u_x}^T v_c) \frac{\partial}{\partial u_w} {u_x}^T v_c$$
$$= \frac{\exp({u_w}^T v_c)}{\sum_{w \in Vocab} \exp({u_w}^T v_c)} v_c = \hat{y}_w v_c$$