Written :

a) As we're dealing with hot vectors, just $y_o = 1$ and other elements are zeros so :

$$-\sum_{w \in V} y_w \log(\hat{y}_w) = -\log(\hat{y}_o)$$

b) $J_{naive\text{-}softmax}(v_c, o, U) = -\log \dfrac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T . v_c)}$

$$\frac{\partial}{\partial v_c} -\log \frac{\exp(u_o^T v_c)}{\sum_w \exp(u_w^T v_c)}$$

$$\frac{\partial}{\partial v_c} \underbrace{-\log(\exp(u_o^T v_c))}_{\frac{\partial}{\partial v_c} -u_o^T v_c = -u_o} + \frac{\partial}{\partial v_c} \underbrace{\log \sum \exp(u_w^T . v_c)}_{\text{partial}}$$

$$\frac{\partial}{\partial v_c} \log \sum_w \exp(u_o^T v_c) = \frac{1}{\sum_{w} \exp(u_w^T v_c)} \cdot \sum_w \underbrace{\frac{\partial}{\partial v_c} \exp(u_w^T v_c)}_{\text{partial}}$$

$$= \frac{1}{\sum_w \exp(u_w^T v_c)} \cdot \sum_w \exp(u_w^T v_c) \cdot \underbrace{\frac{\partial}{\partial v_c} u_w^T v_c}_{u_w}$$

$$= -u_o + \sum_w \frac{\exp(u_w^T v_c)}{\sum_w \exp(u_w^T v_c)} \cdot u_w$$

$$= \underbrace{-u_o}_{-U \cdot y} + \sum_w \underbrace{P(O = w / C = c)}_{U \cdot \hat{y}} . u_w = U(\hat{y} - y)$$

c) $w \neq o$

$$\frac{dJ}{du_w} = -\frac{d}{du_w}(u_o^T v_c) + \frac{d}{du_w}\log\left(\sum_w \exp(u_w^T v_c)\right)$$

0

$$= \sum_w \frac{\exp(u_w^T v_c)}{\sum_w \exp(u_w^T v_c)} \cdot v_c = P(O=w \mid C=c) \cdot v_c$$

---

d) $$\sigma(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1}$$

$$\frac{d\sigma(x)}{dx} = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right) = \sigma(x)(1-\sigma(x))$$

---

e) $$J_{neg\text{-}sampling}(v_c, o, U) = -\log(\sigma(u_o^T v_c)) - \sum_K \log(\sigma(-u_K^T v_c))$$

$$\frac{dJ}{dv_c} = -u_o(1-\sigma(u_o^T v_c)) + \sum_K u_K (1-\sigma(-u_K^T v_c))$$

$$\frac{dJ}{du_o} = -v_c(1-\sigma(u_o^T v_c))$$

$$\frac{dJ}{du_K} = v_c(1-\sigma(-u_K^T v_c)) = (1-\sigma(u_K^T v_c))\, v_c$$

---

f) $$J_{skip\text{-}gram}(v_c, w_{t-m}, \ldots, w_{t+m}, U) = \sum_{-m \le j \le m} J(v_c, w_{t+j}, U)$$

$$\frac{dJ}{dU} = \sum_{-m \le j \le m} \frac{dJ(v_c, w_{t+j}, U)}{dU}$$

$$\frac{dJ}{dv_c} = \sum_{-m \le j \le m} \frac{dJ(v_c, w_{t+j}, U)}{dv_c}$$

$$\frac{dJ}{dv_w} = \sum_{-m \le j \le m} \frac{dJ(v_c, w_{t+j}, U)}{dv_w} = 0$$