(a) Since $y_w = 1$ iff $w = o$, the equation is obvious.

(b) According to the chain rule, the partial derivative is:

$$\frac{\partial \boldsymbol{J}}{\partial v_c} = \frac{\partial_{v_c} P}{P}$$

and we only need to calculate the partial derivative of the i^{th} entry of v_c, which is

$$\partial_i P = \frac{u_{oi} exp(\boldsymbol{u_o^t v_c}) \sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c}) - exp(\boldsymbol{u_o^t v_c}) \sum_{w \in Vocab} u_{wi} exp(\boldsymbol{u_w^t v_c})}{(\sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c}))^2}$$

so

$$\frac{\partial_i P}{P} = \frac{u_{oi} \sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c}) - \sum_{w \in Vocab} u_{wi} exp(\boldsymbol{u_w^t v_c})}{\sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c})}$$

and the partial derivative to $\boldsymbol{v_c}$ should be the concatenation of all $\partial_i P$ in the i^{th} entry:

$$\begin{aligned} \frac{\partial \boldsymbol{J}}{\partial v_c} &= \boldsymbol{u_o} - \frac{\sum_{w \in Vocab} \boldsymbol{u_w} exp(\boldsymbol{u_w^t v_c})}{\sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c})} \\ &= \boldsymbol{U\hat{y}} - \boldsymbol{Uy} \end{aligned}$$

(c) If $w = o$, then

$$\partial_i P = \frac{v_{ci} exp(\boldsymbol{u_o^t v_c}) \sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c}) - exp(\boldsymbol{u_o^t v_c}) v_{ci} exp(\boldsymbol{u_o^t v_c})}{(\sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c}))^2}$$

and use the same method as (b):

$$\frac{\partial_i P}{P} = v_{ci} - \frac{v_{ci} exp(\boldsymbol{u_o^t v_c})}{\sum_{w \in Vocab} exp(\boldsymbol{u_w^t v_c})}$$

then finally:

$$\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u_o}} = (\boldsymbol{\hat{y}}^t \boldsymbol{y} - 1) \boldsymbol{v_c}$$

if $w \neq o$:

$$\partial_i P = \frac{-exp(\boldsymbol{u_o}^t \boldsymbol{v_c}) v_{ci} exp(\boldsymbol{u_w}^t \boldsymbol{v_c})}{(\sum_{w \in Vocab} exp(\boldsymbol{u_w}^t \boldsymbol{v_c}))^2}$$

then

$$\frac{\partial_i P}{P} = \frac{v_{ci} exp(\boldsymbol{u_w}^t \boldsymbol{v_c})}{\sum_{w \in Vocab} exp(\boldsymbol{u_w}^t \boldsymbol{v_c})}$$

$$\begin{aligned} \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u_w}} &= \frac{\boldsymbol{v_c} exp(\boldsymbol{u_w}^t \boldsymbol{v_c})}{\sum_{w \in Vocab} exp(\boldsymbol{u_w}^t \boldsymbol{v_c})} \\ &= \boldsymbol{v_c}(\hat{\boldsymbol{y}}^t \boldsymbol{y}) \end{aligned}$$

so finally

$$\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{U}} = \boldsymbol{v_c}(\hat{\boldsymbol{y}} - \boldsymbol{y})^t$$

(d) It is easy by entry-wise derivative:

$$\boldsymbol{\sigma}'(\boldsymbol{x}) = \boldsymbol{\sigma}(\boldsymbol{x})(1 - \boldsymbol{\sigma}(\boldsymbol{x}))$$

and the multiplication here is also entry-wise other than inner product.