(a).

$$
\begin{aligned}
-\sum_{w \in Vocab} y_w \log(\hat{y}_w) &= -\left[y_o \log(\hat{y}_o) + \sum_{\substack{w \in Vocab \\ w \neq o}} y_w \log(\hat{y}_w) \right] \\
&= -\left[1 \cdot \log(\hat{y}_o) + \sum_{\substack{w \in Vocab \\ w \neq o}} 0 \cdot \log(\hat{y}_w) \right] \\
&= -\log(\hat{y}_o)
\end{aligned}
$$

(b).

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{v}_c} &= \frac{\partial(-\log(\hat{y}_o))}{\partial \boldsymbol{v}_c} \\
&= -\frac{\partial(\log \frac{\exp(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c)}{\sum_{w \in Vocab} \exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c)})}{\partial \boldsymbol{v}_c} \\
&= -\frac{\partial(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c) - \partial(\log \sum_{w \in Vocab} \exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c))}{\partial \boldsymbol{v}_c} \\
&= -\boldsymbol{u}_o + \frac{1}{\sum_{w \in Vocab} \exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c)} \cdot \sum_{w \in Vocab} \boldsymbol{u}_w \exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c) \\
&= -\boldsymbol{u}_o + \sum_{w \in Vocab} \frac{\exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c)}{\sum_{k \in Vocab} \exp(\boldsymbol{u}_k^\intercal \boldsymbol{v}_c)} \cdot \boldsymbol{u}_w \\
&= -\boldsymbol{u}_o + \sum_{w \in Vocab} p(\boldsymbol{u}_w | \boldsymbol{v}_c) \cdot \boldsymbol{u}_w \\
&= -\boldsymbol{u}_o + \sum_{w \in Vocab} \hat{y}_w \boldsymbol{u}_w \\
&= -\boldsymbol{u}_o + \boldsymbol{U}\hat{\boldsymbol{y}} \\
&= -\boldsymbol{U}\boldsymbol{y} + \boldsymbol{U}\hat{\boldsymbol{y}} \\
&= \boldsymbol{U}(\hat{\boldsymbol{y}} - \boldsymbol{y})
\end{aligned}
$$

(c).

case1: $w \neq o$:

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_w} &= \frac{\partial(-\log(\hat{y}_o))}{\partial \boldsymbol{u}_w} \\
&= -\frac{\partial(\log \frac{\exp(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c)}{\sum_{w\in Vocab}\exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c)})}{\partial \boldsymbol{u}_w} \\
&= -\frac{\partial(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c) - \partial(\log \sum_{k\in Vocab}\exp(\boldsymbol{u}_k^\intercal \boldsymbol{v}_c))}{\partial \boldsymbol{u}_w} \\
&= \frac{1}{\sum_{k\in Vocab}\exp(\boldsymbol{u}_k^\intercal \boldsymbol{v}_c)} \cdot \boldsymbol{v}_c \exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c) \\
&= \boldsymbol{v}_c \cdot p(\boldsymbol{u}_w|\boldsymbol{v}_c) \\
&= \boldsymbol{v}_c \cdot \hat{y}_w
\end{aligned}
$$

case 2: $w = o$:

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_w} &= \frac{\partial(-\log(\hat{y}_o))}{\partial \boldsymbol{u}_o} \\
&= -\frac{\partial(\log \frac{\exp(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c)}{\sum_{w\in Vocab}\exp(\boldsymbol{u}_w^\intercal \boldsymbol{v}_c)})}{\partial \boldsymbol{u}_o} \\
&= -\frac{\partial(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c) - \partial(\log \sum_{k\in Vocab}\exp(\boldsymbol{u}_k^\intercal \boldsymbol{v}_c))}{\partial \boldsymbol{u}_o} \\
&= -\boldsymbol{v}_c + \frac{1}{\sum_{k\in Vocab}\exp(\boldsymbol{u}_k^\intercal \boldsymbol{v}_c)} \cdot \boldsymbol{v}_c \exp(\boldsymbol{u}_o^\intercal \boldsymbol{v}_c) \\
&= \boldsymbol{v}_c\,(p(\boldsymbol{u}_o|\boldsymbol{v}_c) - 1) \\
&= \boldsymbol{v}_c \cdot (\hat{y}_o - 1)
\end{aligned}
$$

(d).

$$
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{U}} = \left[\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_1}, \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_2}, \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_3}, ..., \frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_{|Vocab|}}\right]
$$

(e).

$$
\begin{aligned}
\frac{\partial \sigma(x)}{\partial x} &= \frac{\partial(\frac{e^x}{e^x+1})}{\partial x} \\
&= \frac{(e^x)'(e^x+1) - e^x(e^x+1)'}{(e^x+1)^2} \\
&= \frac{e^x(e^x+1) - (e^x)^2}{(e^x+1)^2} \\
&= \sigma(x) - \sigma(x)^2
\end{aligned}
$$

(f).

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{v}_c} &= \frac{\partial(-\log\left(\sigma\left(\boldsymbol{u}_o^\top \boldsymbol{v}_c\right)\right) - \sum_{k=1}^K \log\left(\sigma\left(-\boldsymbol{u}_k^\top \boldsymbol{v}_c\right)\right))}{\partial \boldsymbol{v}_c} \\
&= \frac{\partial}{\partial \boldsymbol{v}_c}\left[-\log(\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c))\right] - \frac{\partial}{\partial \boldsymbol{v}_c}\left[\sum_{k=1}^K \log(\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c))\right] \\
&= \left[-\frac{1}{\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}\left(\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c) - \sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c)^2\right)\boldsymbol{u}_o\right] - \\
&\left[\sum_{k=1}^K \frac{1}{\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c)}\left(\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c) - \sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c)^2\right)(-\boldsymbol{u}_k)\right] \\
&= -\boldsymbol{u}_o\left(1 - \sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c)\right) + \sum_{k=1}^K \boldsymbol{u}_k\left(1 - \sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_o} &= \frac{\partial(-\log\left(\sigma\left(\boldsymbol{u}_o^\top \boldsymbol{v}_c\right)\right) - \sum_{k=1}^K \log\left(\sigma\left(-\boldsymbol{u}_k^\top \boldsymbol{v}_c\right)\right))}{\partial \boldsymbol{u}_o} \\
&= \frac{\partial}{\partial \boldsymbol{u}_o}\left[-\log(\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c))\right] - \frac{\partial}{\partial \boldsymbol{u}_o}\left[\sum_{k=1}^K \log(\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c))\right] \\
&= \left[-\frac{1}{\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}\left(\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c) - \sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c)^2\right)\boldsymbol{v}_c\right] - 0 \\
&= -\boldsymbol{v}_c\left(1 - \sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c)\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_k} &= \frac{\partial(-\log\left(\sigma\left(\boldsymbol{u}_o^\top \boldsymbol{v}_c\right)\right) - \sum_{k=1}^K \log\left(\sigma\left(-\boldsymbol{u}_k^\top \boldsymbol{v}_c\right)\right))}{\partial \boldsymbol{u}_k} \\
&= \frac{\partial}{\partial \boldsymbol{u}_k}\left[-\log(\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c))\right] - \frac{\partial}{\partial \boldsymbol{u}_k}\left[\sum_{t=1}^K \log(\sigma(-\boldsymbol{u}_t^\top \boldsymbol{v}_c))\right] \\
&= 0 - \left[\frac{1}{\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c)}\left(\sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c) - \sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c)^2\right)(-\boldsymbol{v}_c)\right] \\
&= \boldsymbol{v}_c\left(1 - \sigma(-\boldsymbol{u}_k^\top \boldsymbol{v}_c)\right)
\end{aligned}
$$

(g).

$$
\begin{aligned}
\frac{\partial \boldsymbol{J}}{\partial \boldsymbol{u}_k} &= \frac{\partial(-\log\left(\sigma\left(\boldsymbol{u}_o^\top \boldsymbol{v}_c\right)\right) - \sum_{k=1}^{K} \log\left(\sigma\left(-\boldsymbol{u}_k^\top \boldsymbol{v}_c\right)\right))}{\partial \boldsymbol{u}_k} \\
&= \frac{\partial}{\partial \boldsymbol{u}_k}\left[-\log(\sigma(\boldsymbol{u}_o^\top \boldsymbol{v}_c))\right] - \frac{\partial}{\partial \boldsymbol{u}_k}\left[\sum_{t=1}^{K} \log(\sigma(-\boldsymbol{u}_t^\top \boldsymbol{v}_c))\right] \\
&= 0 - \left[\sum_{\boldsymbol{u}_t=\boldsymbol{u}_k} \frac{1}{\sigma(-\boldsymbol{u}_t^\top \boldsymbol{v}_c)}\left(\sigma(-\boldsymbol{u}_t^\top \boldsymbol{v}_c) - \sigma(-\boldsymbol{u}_t^\top \boldsymbol{v}_c)^2\right)(-\boldsymbol{v}_c)\right] \\
&= \sum_{\boldsymbol{u}_t=\boldsymbol{u}_k} \boldsymbol{v}_c\left(1 - \sigma(-\boldsymbol{u}_t^\top \boldsymbol{v}_c)\right)
\end{aligned}
$$

(h).

$$
\partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t-m}, \ldots w_{t+m}, \boldsymbol{U}\right)/\partial \boldsymbol{U} = \sum_{\substack{-m \leq j \leq m \\ j \neq 0}} \partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U}\right)/\partial \boldsymbol{U}
$$

$$
\partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t-m}, \ldots w_{t+m}, \boldsymbol{U}\right)/\partial \boldsymbol{v}_c = \sum_{\substack{-m \leq j \leq m \\ j \neq 0}} \partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U}\right)/\partial \boldsymbol{v}_c
$$

$w \neq c$:

$$
\partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t-m}, \ldots w_{t+m}, \boldsymbol{U}\right)/\partial \boldsymbol{v}_w = \sum_{\substack{-m \leq j \leq m \\ j \neq 0}} \partial \boldsymbol{J}_{\text{skip-gram}}\left(\boldsymbol{v}_c, w_{t+j}, \boldsymbol{U}\right)/\partial \boldsymbol{v}_w
$$