$$CE(\boldsymbol{y}, \hat{\boldsymbol{y}}) = -\boldsymbol{y}^\top \cdot \ln(\hat{\boldsymbol{y}}) = -\sum_i^V y_i \ln(\hat{y}_i) \tag{1}$$

$$\hat{y}_o = p(o|c) = \frac{\exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \tag{2}$$

$$\begin{aligned}
\frac{\partial \ln \hat{y}_o}{\partial \boldsymbol{v}_c} &= \frac{\partial \ln \frac{\exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}}{\partial \boldsymbol{v}_c} \\
&= \frac{\partial \ln \exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} - \frac{\partial \ln \sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} \\
&= \frac{\partial \boldsymbol{u}_o^\top \boldsymbol{v}_c}{\partial \boldsymbol{v}_c} - \frac{\partial \ln \sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} \\
&= \boldsymbol{u}_o - \frac{\partial \ln \sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} \\
&= \boldsymbol{u}_o - \frac{1}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \frac{\partial \sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} \\
&= \boldsymbol{u}_o - \frac{1}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \sum_{w=1}^V \frac{\partial \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} \\
&= \boldsymbol{u}_o - \frac{1}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c) \frac{\partial (\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{v}_c} \\
&= \boldsymbol{u}_o - \frac{1}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c) \boldsymbol{u}_w \\
&= \boldsymbol{u}_o - \sum_{w=1}^V \frac{\exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\sum_{w=1}^V \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \boldsymbol{u}_w \\
&= \boldsymbol{u}_o - \sum_{w=1}^V p(w|c) \cdot \boldsymbol{u}_w \\
&= \boldsymbol{u}_o - \sum_{w=1}^V \hat{y}_w \boldsymbol{u}_w
\end{aligned} \tag{3}$$

$$
\begin{aligned}
\frac{\partial CE(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{v}_c} &= \frac{\partial - \sum_i^V y_i \ln(\hat{y}_i)}{\boldsymbol{v}_c} \\
&= -\frac{\partial \sum_i^V y_i \ln(\hat{y}_i)}{\boldsymbol{v}_c} \\
&= -\frac{\partial \ln(\hat{y}_o)}{\boldsymbol{v}_c} \text{ assume } y_o \text{ is the true class} \\
&= -\boldsymbol{u}_o + \sum_{w=1}^{V} \hat{y}_w \boldsymbol{u}_w \\
&= -\boldsymbol{u}_o + \boldsymbol{U} \cdot \hat{\boldsymbol{y}} \\
&= \boldsymbol{U} \cdot (\hat{\boldsymbol{y}} - \boldsymbol{y}) \\
&\text{To be noticed, } \boldsymbol{y} \text{ and } \hat{\boldsymbol{y}} \text{ are column vectors,} \\
&\text{and } \boldsymbol{U} = [u_1 ... u_V] \text{ in which } u(s) \text{ are column vectors also}
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
\frac{\partial \ln \hat{y}_o}{\partial \boldsymbol{u}_o} &= \frac{\partial \ln \frac{\exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}}{\partial \boldsymbol{u}_o} \\
&= \frac{\partial \ln \exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_o} - \frac{\partial \ln \sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_o} \\
&= \boldsymbol{v}_c - \frac{\partial \ln \sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_o} \\
&= \boldsymbol{v}_c - \frac{1}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \frac{\partial \sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_o} \\
&= \boldsymbol{v}_c - \frac{1}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \frac{\partial \exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_o} \\
&= \boldsymbol{v}_c - \frac{1}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c) \cdot \frac{\partial (\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_o} \\
&= \boldsymbol{v}_c - \frac{\exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \boldsymbol{v}_c \\
&= \boldsymbol{v}_c - p(o|c) \cdot \boldsymbol{v}_c \\
&= (1 - \hat{y}_o) \cdot \boldsymbol{v}_c
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
\frac{\partial \ln \hat{y}_o}{\partial \boldsymbol{u}_x} &= \frac{\partial \ln \frac{\exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}}{\partial \boldsymbol{u}_x} \\
&= \frac{\partial \ln \exp(\boldsymbol{u}_o^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_x} - \frac{\partial \ln \sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_x} \\
&= -\frac{\partial \ln \sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_x} \\
&= -\frac{1}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \frac{\partial \sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_x} \\
&= -\frac{1}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \frac{\partial \exp(\boldsymbol{u}_x^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_x} \\
&= -\frac{1}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \exp(\boldsymbol{u}_x^\top \boldsymbol{v}_c) \cdot \frac{\partial (\boldsymbol{u}_x^\top \boldsymbol{v}_c)}{\partial \boldsymbol{u}_x} \\
&= -\frac{\exp(\boldsymbol{u}_x^\top \boldsymbol{v}_c)}{\sum_{w=1}^{V} \exp(\boldsymbol{u}_w^\top \boldsymbol{v}_c)} \cdot \boldsymbol{v}_c \\
&= -p(x|c) \cdot \boldsymbol{v}_c \\
&= -\hat{y}_x \cdot \boldsymbol{v}_c
\end{aligned} \tag{6}
$$

$$
\begin{aligned}
\frac{\partial CE(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{U}} &= \frac{\partial - \sum_i^V y_i \ln(\hat{y}_i)}{\boldsymbol{U}} \\
&= -\frac{\partial \sum_i^V y_i \ln(\hat{y}_i)}{\boldsymbol{U}} \\
&= -\frac{\partial \ln(\hat{y}_o)}{\boldsymbol{U}} \text{ assume } y_o \text{ is the true class} \\
&= -\frac{\partial \ln(\hat{y}_o)}{\boldsymbol{U}}
\end{aligned} \tag{7}
$$

$$
\begin{aligned}
\frac{\partial CE(\boldsymbol{y}, \hat{\boldsymbol{y}})}{\partial \boldsymbol{U}} &= \begin{cases} = (\hat{y}_o - 1) \cdot \boldsymbol{v}_c, & w = o \\ = \hat{y}_x \cdot \boldsymbol{v}_c, & w \neq o \end{cases} \\
&= \boldsymbol{v}_c \cdot (\hat{\boldsymbol{y}} - \boldsymbol{y})^\top \\
&\text{To be noticed, } \boldsymbol{v}_c, \boldsymbol{y} \text{ and } \hat{\boldsymbol{y}} \text{ are column vectors} \\
&\text{In numpy you should use np.outer}
\end{aligned} \tag{8}
$$