2(b) $\dfrac{\partial(CE(\mathbf{y},\hat{\mathbf{y}}))}{\partial \boldsymbol{\theta}}$

$$= -\frac{\partial}{\partial \boldsymbol{\theta}} \sum_i \left(y_i \log(y_i) \right) \tag{1}$$

$$= \begin{bmatrix} -\dfrac{\partial}{\partial \theta_1}\left(\left(y_1 \log y_1\right) + \left(y_2 \log y_2\right) + \ldots + \left(y_k \log y_k\right) + \ldots\right) \\ -\dfrac{\partial}{\partial \theta_2}\left(\left(y_1 \log y_1\right) + \left(y_2 \log y_2\right) + \ldots + \left(y_k \log y_k\right) + \ldots\right) \\ \vdots \\ -\dfrac{\partial}{\partial \theta_k}\left(\left(y_1 \log y_1\right) + \left(y_2 \log y_2\right) + \ldots + \left(y_k \log y_k\right) + \ldots\right) \\ \vdots \end{bmatrix} \tag{2}$$

$$= \begin{bmatrix} -\dfrac{\partial}{\partial \theta_1}\left(\left(y_1 \log\left(\dfrac{\exp(\theta_1)}{\sum_m \exp(\theta_m)}\right)\right) + \left(y_2 \log\left(\dfrac{\exp(\theta_2)}{\sum_m \exp(\theta_m)}\right)\right) + \ldots + \left(y_k \log\left(\dfrac{\exp(\theta_k)}{\sum_m \exp(\theta_m)}\right)\right) + \ldots\right) \\ -\dfrac{\partial}{\partial \theta_2}\left(\left(y_1 \log\left(\dfrac{\exp(\theta_1)}{\sum_m \exp(\theta_m)}\right)\right) + \left(y_2 \log\left(\dfrac{\exp(\theta_2)}{\sum_m \exp(\theta_m)}\right)\right) + \ldots + \left(y_k \log\left(\dfrac{\exp(\theta_k)}{\sum_m \exp(\theta_m)}\right)\right) + \ldots\right) \\ \vdots \\ -\dfrac{\partial}{\partial \theta_k}\left(\left(y_1 \log\left(\dfrac{\exp(\theta_1)}{\sum_m \exp(\theta_m)}\right)\right) + \left(y_2 \log\left(\dfrac{\exp(\theta_2)}{\sum_m \exp(\theta_m)}\right)\right) + \ldots + \left(y_k \log\left(\dfrac{\exp(\theta_k)}{\sum_m \exp(\theta_m)}\right)\right) + \ldots\right) \\ \vdots \end{bmatrix} \tag{3}$$

If we look at the j^{th} row of Equation (3), it is of the form below:

$$= -\frac{\partial}{\partial \theta_j}\left(\sum_i y_i \log\left(\frac{\exp(\theta_i)}{\sum_m \exp(\theta_m)}\right)\right) \tag{4}$$

$$= -\sum_i \frac{\partial}{\partial \theta_j}\left(y_i \log\left(\frac{\exp(\theta_i)}{\sum_m \exp(\theta_m)}\right)\right) \tag{5}$$

=	$-\sum_i \frac{\partial}{\partial\theta_j}\left(y_i\left(\log\left(\exp(\theta_i)\right)-\log\sum_m \exp(\theta_m)\right)\right)$	(6)
=	$-\sum_i y_i\left(\frac{\partial\theta_i}{\partial\theta_j}-\frac{\partial}{\partial\theta_j}\left(\log\sum_m \exp(\theta_m)\right)\right)$	(7)
=	$-\sum_i y_i\left(\frac{\partial\theta_i}{\partial\theta_j}-\left(\frac{1}{\sum_m \exp(\theta_m)}\right)\frac{\partial}{\partial\theta_j}\left(\sum_x \exp(\theta_x)\right)\right)$	(8)
=	$-\sum_i y_i\left(\frac{\partial\theta_i}{\partial\theta_j}-\sum_x\frac{\frac{\partial}{\partial\theta_j}\left(\exp(\theta_x)\right)}{\sum_m \exp(\theta_m)}\right)$	(9)

When $i = j$ in Equation (9):

=	$-\left(\ldots+y_j\left(\frac{\partial\theta_j}{\partial\theta_j}-\frac{\frac{\partial}{\partial\theta_j}\left(\exp(\theta_1)+\exp(\theta_2)+\ldots.+\exp(\theta_j)+\ldots\right)}{\sum_m \exp(\theta_m)}\right)+\ldots\right)$	(10)
=	$-\left(\ldots+y_j\left(1-\frac{\exp(\theta_j)}{\sum_m \exp(\theta_m)}\right)+\ldots\right)$	(11)
=	$-\left(\ldots+y_j\left(1-y_j\right)+\ldots\right)$	(12)
=	$-\left(\ldots+y_j-y_j y_j+\ldots\right)$	(13)

Now let's consider all cases where $i \neq j$ in Equation(9):

=	$-\left(\ldots+y_{i\neq j}\left(\frac{\partial\theta_{i\neq j}}{\partial\theta_j}-\frac{\frac{\partial}{\partial\theta_j}\left(\exp(\theta_1)+\exp(\theta_2)+\ldots.+\exp(\theta_j)+\ldots\right)}{\sum_m \exp(\theta_m)}\right)+\ldots\right)$	(14)
=	$-\left(\ldots+y_{i\neq j}\left(0-\frac{\exp(\theta_j)}{\sum_m \exp(\theta_m)}\right)+\ldots\right)$	(15)

$$= -\left(\ldots + y_{i \neq j}\left(0 - y_j\right) + \ldots\right) \tag{16}$$

$$= -\left(\ldots + -y_{i \neq j} y_j + \ldots\right) \tag{17}$$

Combining Equations (13) and (17):

$$= -\left(y_j - y_j y_j - y_{i \neq j} y_j + \ldots\right) \tag{18}$$

$$= -y_j + y_j y_j + y_{i \neq j} y_j + \ldots \tag{19}$$

$$= -y_j + y_j \sum_n y_n \tag{20}$$

Now, $\sum_n y_n = 1$ because $\mathbf{y}$ is a one-hot label vector whose elements are 0 except for the k^{th} dimension (only $y_k = 1$). Substituting in Equation (20):

$$= -y_j + y_j(1) \tag{21}$$

$$= y_j - y_j \tag{22}$$

Substituting in Equation (3):

$$\frac{\partial(CE(\mathbf{y}, \hat{\mathbf{y}}))}{\partial \boldsymbol{\theta}} = \begin{bmatrix} y_1 - y_1 \\ y_2 - y_2 \\ \vdots \\ y_k - y_k \\ \vdots \end{bmatrix} \tag{23}$$

$$= \hat{\mathbf{y}} - \mathbf{y} \tag{24}$$

Assuming k is the correct class (i.e., $y_k = 1$):

$$= \begin{cases} y_k - 1, & i = k \\ y_k, & \text{otherwise} \end{cases} \tag{25}$$