Exercise 1

Task 1:

$y_i = \hat{y}_i \quad z_j = \tilde{z}_j^{(L)}$

$$\frac{\partial L}{\partial \tilde{z}^{(L)}} = \sum_{i=1}^{C} \frac{\partial L}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial \tilde{z}_j^{(L)}} \quad \text{(chain rule)}$$

$$L = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \Rightarrow \frac{\partial L}{\partial \hat{y}_i} = -y_i \cdot \frac{1}{\hat{y}_i}$$

$$\frac{\partial \hat{y}_i}{\partial \tilde{z}_j^{(L)}} = \hat{y}_i (\delta_{ij} - \hat{y}_j)$$

$$\Rightarrow \frac{\partial L}{\partial \tilde{z}^{(L)}} = \sum^{C} \left(-\frac{y_i}{\hat{y}_i}\right) \cdot \hat{y}_i (\delta_{ij} - \hat{y}_j) = \sum^{C} -y_i (\delta_{ij} - \hat{y}_j)$$

$$= -y_j + \hat{y}_j \underbrace{\sum^{C} y_j}_{=1} = \hat{y}_j - y_j$$

Task 2:

a)

$$\frac{\partial L}{\partial W^{(L-1)}} = \frac{\partial L}{\partial \tilde{z}^{(L)}} \cdot \frac{\partial \tilde{z}^{(L)}}{\partial W^{(L-1)}}$$

$$\nabla_{W^{(L-1)}} L = \frac{\partial L}{\partial \tilde{z}^{(L)}} = \hat{y} - y := \delta^{(L)} \in \mathbb{R}^C$$

b) $\cancel{\frac{\partial \tilde{z}^{(L)}}{\partial W^{(L-1)}} = \nabla_{b^{(L-1)}} L}$

$$\tilde{z}^{(L)} = \sum_{j=1}^{d_{L-1}} w_{ij}^{(L-1)} \cdot z_j^{(L-1)} + b_i^{(L-1)}$$

$$\frac{\partial L}{\partial W_{ij}^{(L-1)}} = \delta_i^{(L)} \cdot z_j^{(L-1)}$$

Task 2

b)

$$\frac{\partial L}{\partial_b (L-1)} = \nabla_{b^{(L-1)}} L = \frac{\partial L}{\partial \tilde{z}^{(L)}} \cdot \underbrace{\frac{\partial \tilde{z}^{(L)}}{\partial b^{(L-1)}}}_{=1} = \delta^{(L)} = \hat{y} - y$$

Because

$$\tilde{z}^{(L)} = W^{(L-1)} z^{(L-1)} + b^{(L-1)}$$

$$\Rightarrow \tilde{z}_i^{(L)} = \underbrace{\sum_{j=1}^{d_{L-1}} W_{ij}^{(L-1)} z_j^{(L-1)} + b_i^{(L-1)}}_{\text{Linear}} \quad \Big| \quad \frac{\partial \tilde{z}_i^{(L)}}{\partial b_i^{(L-1)}} = 1$$

Dimensions:

a) $\delta(L) \in \mathbb{R}^c$ $\Rightarrow$

$W^{(L-1)} \in \mathbb{R}^{c \times d_{L-1}}$ ~~Because~~ Because $z^{(L-1)} \in \mathbb{R}^{d_{L-1}}$

b) $b^{(L-1)} \in \mathbb{R}^C$

$\delta^{(L)} \in \mathbb{R}^C$

Task 3 $\delta^{(L-1)} = ((W^{(L-1)})^T \, \delta^{(L)}) \odot \phi'(\tilde{z}^{(L-1)})$

$\phi(x) = \max(0, x)$ $\quad \phi'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{else} \end{cases}$

When $\tilde{z}_j^{(L-1)} < 0$:

$\rightarrow \phi'(\tilde{z}^{(L-1)}) = 0$

$\rightarrow \delta_j^{(L-1)} = 0$

$\Rightarrow$ If the neuron does not activate, then it does not contribute to the learning