QUESTION 4:

a) $\text{softmax}(x) = \text{softmax}(x+c)$

$$\text{softmax}_i(x+c) = \frac{e^{x_i+c}}{\sum_{j=1}^{N} e^{x_j+c}} = \frac{\cancel{e^{c}}\, e^{x_i}}{\cancel{e^{c}} \sum_{j=1}^{N} e^{x_j}}$$

$$= \text{softmax}_i(x)$$

b) $D_x \log(\sigma(x)) \in \mathbb{R}^{N \times N}$

$\{\log = \ln\}$

$$\log \sigma_i(x) = \log \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} = \log e^{x_i} - \log\left(\sum_{j=1}^{N} e^{x_j}\right)$$

$$= x_i - \log\left(\sum_{j=1}^{N} e^{x_j}\right)$$

diagonal:

$$\frac{\partial}{\partial x_i} \log(\sigma_i(x)) = 1 - \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \star = \boxed{1 - \sigma_i(x)}$$

$$\star \quad \frac{\partial \log(f(x))}{\partial x} = \frac{1}{f(x)} \cdot f'(x)$$

off-diagonal:

$$\frac{\partial}{\partial x_j} \log(\sigma_i(x)) = \cancel{\frac{\partial x_i}{\partial x_j}}^{\,0} - \frac{\partial}{\partial x_j} \log\left(\sum_{k=1}^{N} e^{x_k}\right)$$

$$= \frac{-e^{x_j}}{\sum_{k=1}^{N} e^{x_k}} = \boxed{-\sigma_j(x)}$$