N = number of classes

$x \in \mathbb{R}^{1 \times N}$: predicted scores for classes

$x \in \mathbb{R}\ [-\infty, \infty]$

softmax: $\sigma(x) = \dfrac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \Rightarrow \sum_{i=1}^{N} \sigma_i(x) = 1$, ~~$\sigma_i(x) \in$~~

$\sigma_i(x) \in [0, 1] \ \forall i$

QUESTION 1:

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

$$\begin{bmatrix} \dfrac{\partial \sigma_1(x)}{\partial x_1} & \cdots & \dfrac{\partial \sigma_1(x)}{\partial x_N} \\ \vdots & \ddots & \\ \dfrac{\partial \sigma_N(x)}{\partial x_1} & & \dfrac{\partial \sigma_N(x)}{\partial x_N} \end{bmatrix}$$

diagonal: quotient rule $\left(\dfrac{u}{v}\right)' = \dfrac{u'v - uv'}{v^2}$

$$u = e^{x_i}, \quad v = \sum_{j=1}^{N} e^{x_j}$$

$$\frac{\partial \sigma_i(x)}{\partial x_i} = \frac{e^{x_i}\left(\sum_{j=1}^{N} e^{x_j}\right) - e^{x_i} e^{x_i}}{\left(\sum_{j=1}^{N} e^{x_j}\right)^2} \quad * \quad \cancel{\sum_{j=1}^{N} e^{x_j}}$$

$$\sum_{j=1}^{N} e^{x_j} = e^{x_i} + \cancel{\sum_{j=1, j \neq i}^{N} e^{x_j}} = -e^{x_i}$$