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

$$= \boxed{\sigma_i(x)\,(1 - \sigma_i(x))}$$

Off-diagonal:

rule: $(cu^{-1})' = -cu^{-2} \cdot u'$ ~~(c = cons~~

(c = constant)

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

$$\frac{\partial \sigma_i(x)}{\partial x_j} = \frac{-e^{x_i} e^{x_j}}{\left(\sum_{k=1}^{N} e^{x_k}\right)^2} \rightarrow \sum_{k=1}^{N} e^{x_k} = \cancel{e^{x_j} + \sum_{k=1, k\neq j}^{N} e^{x_k}}$$

$$= e^{x_j} + \sum_{k=1, k\neq j}^{N} e^{x_k}$$

$$= \boxed{-\sigma_i(x)\sigma_j(x)}$$

Combining:

$$\begin{bmatrix} \sigma_1(x)(1-\sigma_1(x)) & \cdots & -\sigma_N(x)\sigma_1(x) \\ -\sigma_1(x)\sigma_2(x) & \ddots & \vdots \\ \vdots & & \\ -\sigma_1(x)\sigma_N(x) & & \sigma_N(x)(1-\sigma_N(x)) \end{bmatrix}$$

b) $Z = v D_x \sigma(x) = [\underbrace{v_1 \, v_2 \cdots v_N}_{N}] \left[N \times N \right]$

$$Z_1 = v_1 \sigma_1(x)(1-\sigma_1(x)) + v_2(-\sigma_1(x)\sigma_2(x)) + \cdots$$
$$+ \cdots v_N(-\sigma_1(x)\sigma_N(x))$$
$$= \sigma_1(x)\left(v_1 - \cancel{v_1\sigma_1(x)} - \sum_{j=1}^{N} v_j \sigma_j(x) + \cancel{v_1\sigma_1(x)}\right)$$
$$= \sigma_1(x)\left(v_1 - v\,\sigma(x)^T\right)$$