1.e

We start by writing $\varsigma(x)$ with $x \in \mathbb{R}^n$ in its vector form:

$$\varsigma(x) = \begin{bmatrix} \frac{e^{x_1}}{\sum_{j=1}^{n} e^{x_j}} \\ \vdots \\ \frac{e^{x_n}}{\sum_{j=1}^{n} e^{x_j}} \end{bmatrix}$$

Thus, getting the derivative of ς w.r.t. the vector x is just a matter of applying the partial derivatives to each one of the rows.

$$\frac{\partial}{\partial x}\varsigma(x) = \begin{bmatrix} \frac{\partial}{\partial x_1}\frac{e^{x_1}}{\sum_{j=1}^{n} e^{x_j}} & \cdots & \frac{\partial}{\partial x_n}\frac{e^{x_1}}{\sum_{j=1}^{n} e^{x_j}} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1}\frac{e^{x_n}}{\sum_{j=1}^{n} e^{x_j}} & \cdots & \frac{\partial}{\partial x_n}\frac{e^{x_n}}{\sum_{j=1}^{n} e^{x_j}} \end{bmatrix}$$

We then start by computing the derivative of a cell to fill out the matrix afterwards. For x_i we have:

$$\frac{\partial}{\partial x_i}\varsigma(x)_i = \frac{\partial}{\partial x_i}\frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} =$$

$$= \frac{\partial}{\partial x_i}\frac{e^{x_i}}{e^{x_1} + \ldots + e^{x_n}} =$$

Applying the quotient rule to the fraction and then the sum rule to $(e^{x_1} + \ldots + e^{x_n})$, we get:

$$= \frac{e^{x_i}(e^{x_1} + \ldots + e^{x_n}) - e^{x_i}(0 + \ldots + e^{x_i} + \ldots + 0)}{(e^{x_1} + \ldots + e^{x_n})^2} = \frac{e^{x_i}(e^{x_1} + \ldots + e^{x_n}) - e^{2x_i}}{(e^{x_1} + \ldots + e^{x_n})^2} =$$

$$= \frac{e^{x_i}}{(e^{x_1} + \ldots + e^{x_n})} - \frac{e^{x_i}}{(e^{x_1} + \ldots + e^{x_n})}\frac{e^{x_i}}{(e^{x_1} + \ldots + e^{x_n})} = \varsigma(x)_i - \varsigma(x)_i^2$$

Now for x_k with $k \neq i$, we have:

$$\frac{\partial}{\partial x_k}\varsigma(x)_i = \frac{\partial}{\partial x_k}\frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} =$$

$$= \frac{\partial}{\partial x_k}\frac{e^{x_i}}{e^{x_1} + \ldots + e^{x_n}} =$$

Applying the quotient rule to the fraction and then the sum rule to $(e^{x_1} + \ldots + e^{x_n})$, we get:

$$= \frac{0(e^{x_1} + \ldots + e^{x_n}) - e^{x_i}(0 + \ldots + e^{x_k} + \ldots + 0)}{(e^{x_1} + \ldots + e^{x_n})^2} = \frac{-e^{x_i + x_k}}{(e^{x_1} + \ldots + e^{x_n})^2} =$$

$$= -\frac{e^{x_i}}{(e^{x_1} + \ldots + e^{x_n})}\frac{e^{x_k}}{(e^{x_1} + \ldots + e^{x_n})} = -\varsigma(x)_i\varsigma(x)_k$$

With all the matrix entries in hand, we can write the derivative down as $\frac{\partial}{\partial x}\varsigma(x) = \mathrm{diag}(\varsigma(x)) - \varsigma(x)\varsigma(x)^T$.