1.2 Calculus-Q2

The softmax function $S : \mathbb{R}^k \to [0,1]^k$, is defined as follows:

$$S(x)_j = \frac{e^{x_j}}{\sum_{l=1}^{k} e^{x_l}}$$

This function takes an input vector $x \in \mathbb{R}^d$ and outputs a probability vector (non-negative entries that sum up to 1), corresponding to the weight of original entries of x.
CQ2: Calculate the Jacobian of the softmax function S.

$$S'(x) = \begin{bmatrix} \frac{\partial S_1}{\partial x_1} & \cdots & \frac{\partial S_1}{\partial x_K} \\ & & \\ & & \frac{\partial S_K}{\partial x_K} \end{bmatrix}$$

let $Z = \sum_{i=1}^{K} e^{x_i}$.

$$\frac{\partial S_i}{\partial x_j} : \frac{d}{dx_j} \cdot \frac{e^{x_i}}{\sum x_j} \overset{*}{=} \frac{\frac{d}{dx_j}(e^{x_i}) \cdot Z - e^{x_i} \cdot \frac{d}{dx_j}(Z)}{Z^2}$$

$$\overset{**}{=} \frac{\delta_{ij} \cdot e^{x_i} \cdot Z - e^{x_i} e^{x_j}}{Z^2}$$

$$= \frac{e^{x_i}}{Z}\left(\delta_{ij} - \frac{e^{x_j}}{Z}\right)$$

$$* \quad \frac{f(x)}{g(x)} = \frac{f'(x) g(x) - f(x) g'(x)}{g(x)^2}$$

$$\cdot \quad \left(\sum e^{x_j}\right)' = \sum \left(e^{x_j}\right)' = \sum e^{x_j}$$

$$** \quad \frac{d}{dx_j} e^{x_i} = \begin{cases} e^{x_i} & i=j \\ 0 & i \neq j \end{cases}$$

$$\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

$$\frac{d}{dx_j}(Z) = \frac{d}{dx_j} \sum e^{x_i} = \begin{cases} e^{x_j} & i=j \\ 0 & i \neq j \end{cases}$$

therefore

$$J_{ij} = \begin{cases} S(x)_i \cdot (1 - S(x))_j & i=j \\ -\frac{e^{x_i} \cdot e^{x_j}}{Z^2} = -S(x)_i S(x)_j & i \neq j \end{cases}$$

$$\boxed{J = \begin{bmatrix} S(x_1) \cdot (1 - S(x_1)) & -S(x_1) \cdot S(x_2) & \cdots & -S(x_1) S(x_K) \\ \vdots & S(x_2) \cdot (1 - S(x_2)) & & \\ & & \ddots & \\ -S(x_K) \cdot S(x_1) & & & S(x_K) \cdot (1 - S(x_K)) \end{bmatrix}}$$