1 Part A: Calculus

1.1 Calculus-Q1

CQ1.1 Use the chain rule to calculate the gradient of

$$h(\boldsymbol{\sigma}) = \frac{1}{2}\|f(\boldsymbol{\sigma}) - y\|^3$$

where $\boldsymbol{\sigma} \in \mathbb{R}^m$ and f is some arbitrary function from $\mathbb{R}^m$ to $\mathbb{R}^n$.
CQ1.2 Compute the expression in the case where:

$$f(\boldsymbol{\sigma}) = \begin{bmatrix} \sigma_1 \cdot \sigma_2 \\ \sigma_1^2 + \sigma_2^2 \\ \sigma_1 \end{bmatrix}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$h(\sigma) = \frac{1}{2} \cdot \|r(\sigma)\|^3$

let:

$r(\sigma) = f(\sigma) - y, \quad g(u) = \frac{1}{2}\|u\|^3$

hence:

$h(\sigma) = g(r(\sigma))$

let us find gradiants:

$\|u\| = (u \cdot u^T)^{\frac{1}{2}}$

$g(u) = \frac{1}{2}\|u\|^3 = \frac{1}{2}(u \cdot u^T)^{\frac{3}{2}}$

$g'(u) = \frac{1}{2} \cdot \frac{3}{2} \cdot (u \cdot u^T)^{\frac{1}{2}} \cdot (u \cdot u^T)' = \frac{1}{2} \cdot \frac{3}{2} \cdot (u \cdot u^T)^{\frac{1}{2}} \cdot (u^2)'$

$= \frac{1}{2} \cdot \frac{3}{\cancel{2}} \cdot \|u\| \cdot \cancel{2}u = \frac{3}{2}\|u\| \cdot u =$

let $u = r(\sigma)$:

$g'(u) = \frac{3}{2}\|r(\sigma)\| r(\sigma)$

we evaluate $f'(\alpha)$ as the jacobian:

$$J_f(\sigma) = \begin{bmatrix} \frac{\partial f_1}{\partial \sigma_1} & \cdots & \frac{\partial f_1}{\partial \sigma_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial \sigma_1} & \cdots & \frac{\partial f_n}{\partial \sigma_m} \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

$$f'(\sigma) = \begin{pmatrix} 2 & 1 \\ 2 & 4 \\ 1 & 0 \end{pmatrix}$$

we want to find gradiant $h(\sigma)$

evaluated at:

$$f(\boldsymbol{\sigma}) = \begin{bmatrix} \sigma_1 \cdot \sigma_2 \\ \sigma_1^2 + \sigma_2^2 \\ \sigma_1 \end{bmatrix}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

by the chainrule:

$$h'(\sigma) = g'(r(\sigma)) \cdot r'(\sigma)$$

$$h'(\sigma) = g'(r(\sigma)) \cdot f'(\sigma)$$

$$= \frac{3}{2} \|r(\sigma)\| \cdot r(\sigma) \cdot f'(\sigma)$$

$$= \frac{3}{2} \|r(\sigma)\| \cdot J_r(\sigma)^T \cdot r(\sigma)$$

let us evaluate:

$$\sigma = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad y = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$f(\sigma) = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} \rightarrow f'(\sigma) = \begin{pmatrix} 2 & 1 \\ 2 & 4 \\ 1 & 0 \end{pmatrix}$$

$$r(\sigma) = \begin{matrix} 1 \\ 4 \\ 0 \end{matrix}$$

$$h'(\sigma) = \frac{3}{2} \cdot \left\| \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix} \right\| \cdot \begin{pmatrix} 2 & 2 & 1 \\ 1 & 4 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 4 \\ 0 \end{pmatrix}$$

$$\boxed{= \frac{3}{2} \cdot \sqrt{17} \cdot \begin{pmatrix} 10 \\ 17 \end{pmatrix}}$$

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

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

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

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

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

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

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

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

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

therefore

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

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