Question.3-17

Question.3-18에서는 Jacobian들의 matrix multiplication을 이용하여 $y = \theta x$를 학습시키는 과정을 증명하였다.

그리고 이 과정에서 $\frac{\partial \vec{z_2}}{\partial \vec{z_1}}, \frac{\partial \vec{\mathscr{L}}}{\partial \vec{z_2}}$는 모두 diagonal matrix인 것을 확인할 수 있었다.

이 관점에서 다음의 질문들에 답하시오.

1) nxn diagonal matrix A, B의 matrix multiplication은 각 diagonal entry들을 원소로 가지는 vector들의 Hadamard product으로 표현할 수 있음을 보이시오.

2) Question.3-16의 과정을 1)의 내용을 이용하여 Hadamard product으로 표현하시오.

1) 먼저 주어진 $\vec{a}$와 B를 다음과 같이 정의하자. 그리고 B의 diagonal entry를 원소로 하는 column vector $\vec{\beta}$를 정의하면 다음과 같다.

$$\vec{a} = (a_1 \ a_2 \ \cdots \ a_n) \qquad B = \begin{pmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{pmatrix} \Rightarrow \vec{\beta} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

이때 $\vec{a} \cdot B$는

$$\vec{a} \cdot B = (a_1 \ a_2 \ \cdots \ a_n) \begin{pmatrix} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_n \end{pmatrix} = (a_1b_1 \ a_2b_2 \ \cdots \ a_nb_n)$$

이 된다. 그리고 $\vec{a}^T$와 $\vec{\beta}$의 Hadamard product는 다음과 같다.

$$\vec{a}^T \circ \vec{\beta} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \circ \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1b_1 \\ a_2b_2 \\ \vdots \\ a_nb_n \end{pmatrix}$$

위의 결과에서 알 수 있듯이 row vector $\vec{a}$와 diagonal matrix B의 결과에서 0이 아닌 값들은 모두 $\vec{a}^T$와 $\vec{\beta}$의 Hadamard product으로 표현된다.

2) 주어진 조건의 $\vec{a}, \vec{b}$를

$$\vec{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

라 하면 $\vec{a}$와 $\vec{b}$의 dot product는 다음과 같다.

$$\vec{a}^T \cdot \vec{b} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}^T \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i$$

그리고 $\vec{a}$와 $\vec{b}$의 Hadamard product는

$$\vec{a} \circ \vec{b} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \circ \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1b_1 \\ a_2b_2 \\ \vdots \\ a_nb_n \end{pmatrix}$$

이 된다. 따라서 Hadamard product를 계산한 뒤 $\vec{a} \circ \vec{b}$의 결과를 모두 더해주면 $\vec{a}^T \cdot \vec{b}$의 결과와 같다. 즉

$$\vec{a}^T \cdot \vec{b} = \mathrm{sum}(\vec{a} \circ \vec{b})$$

로 표현할 수 있다. 이때 sum( )은 벡터의 원소들의 합을 의미한다.

3) 먼저 Question.3-16에서 Jacobian들을 구한 결과를 보면

$$\frac{\partial \vec{z}_2}{\partial \vec{z}_1} = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix} \qquad \frac{\partial \mathcal{J}}{\partial \vec{\mathcal{L}}} = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix}$$

$$\theta,\ \vec{x} \longrightarrow \boxed{\vec{z}_1 = \theta \cdot \vec{x}} \longrightarrow \boxed{\vec{z}_2 = \vec{y} - \vec{z}_1} \longrightarrow \boxed{\vec{\mathcal{L}} = \vec{z}_2 \circ \vec{z}_2} \longrightarrow \boxed{\mathcal{J} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}}$$

$$\vec{y} \longrightarrow \boxed{\vec{z}_2 = \vec{y} - \vec{z}_1}$$

$$\frac{\partial \vec{z}_1}{\partial \theta} = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{pmatrix} \qquad \frac{\partial \vec{\mathcal{L}}}{\partial \vec{z}_2} = \begin{pmatrix} 2z_2^{(1)} & 0 & \cdots & 0 \\ 0 & 2z_2^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2z_2^{(n)} \end{pmatrix}$$

이다. 여기 Hadamard product를 이용하기 위해 다음과 같이 새로운 vector들을 정의하자.

$$\vec{\gamma} = \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} \quad \frac{\partial \vec{z}_2}{\partial \vec{z}_1} = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix} \qquad \frac{\partial \mathcal{J}}{\partial \vec{\mathcal{L}}} = \begin{pmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{pmatrix} \quad \vec{\alpha} = \begin{pmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix}$$

$$\theta,\ \vec{x} \longrightarrow \boxed{\vec{z}_1 = \theta \cdot \vec{x}} \longrightarrow \boxed{\vec{z}_2 = \vec{y} - \vec{z}_1} \longrightarrow \boxed{\vec{\mathcal{L}} = \vec{z}_2 \circ \vec{z}_2} \longrightarrow \boxed{\mathcal{J} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}^{(i)}}$$

$$\vec{y} \longrightarrow \boxed{\vec{z}_2 = \vec{y} - \vec{z}_1}$$

$$\frac{\partial \vec{z}_1}{\partial \theta} = \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{pmatrix} \qquad \frac{\partial \vec{\mathcal{L}}}{\partial \vec{z}_2} = \begin{pmatrix} 2z_2^{(1)} & 0 & \cdots & 0 \\ 0 & 2z_2^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2z_2^{(n)} \end{pmatrix} \quad \vec{\beta} = \begin{pmatrix} 2z_2^{(1)} \\ 2z_2^{(2)} \\ \vdots \\ 2z_2^{(n)} \end{pmatrix} = \begin{pmatrix} 2(y^{(1)} - \hat{y}^{(1)}) \\ 2(y^{(2)} - \hat{y}^{(2)}) \\ \vdots \\ 2(y^{(n)} - \hat{y}^{(n)}) \end{pmatrix}$$

위의 $\vec{\alpha}, \vec{\beta}, \vec{\gamma}$를 이용하여 기존 연산을 표현하면 다음과 같다.

$$\frac{\partial \mathcal{J}}{\partial \vec{\mathcal{L}}} \cdot \frac{\partial \vec{\mathcal{L}}}{\partial \vec{z}_2} \Rightarrow \vec{\alpha} \circ \vec{\beta} = \begin{pmatrix} \frac{1}{n} \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix} \circ \begin{pmatrix} 2(y^{(1)} - \hat{y}^{(1)}) \\ 2(y^{(2)} - \hat{y}^{(2)}) \\ \vdots \\ 2(y^{(n)} - \hat{y}^{(n)}) \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \cdot 2(y^{(1)} - \hat{y}^{(1)}) \\ \frac{1}{n} \cdot 2(y^{(2)} - \hat{y}^{(2)}) \\ \vdots \\ \frac{1}{n} \cdot 2(y^{(n)} - \hat{y}^{(n)}) \end{pmatrix}$$

$$\frac{\partial \mathcal{J}}{\partial \vec{z}_2} \cdot \frac{\partial \vec{z}_2}{\partial \vec{z}_1} \Rightarrow (\vec{\alpha} \circ \vec{\beta}) \circ \vec{\gamma} = \begin{pmatrix} \frac{1}{n} \cdot 2(y^{(1)} - \hat{y}^{(1)}) \\ \frac{1}{n} \cdot 2(y^{(2)} - \hat{y}^{(2)}) \\ \vdots \\ \frac{1}{n} \cdot 2(y^{(n)} - \hat{y}^{(n)}) \end{pmatrix} \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{n} \cdot (-2(y^{(1)} - \hat{y}^{(1)})) \\ \frac{1}{n} \cdot (-2(y^{(2)} - \hat{y}^{(2)})) \\ \vdots \\ \frac{1}{n} \cdot (-2(y^{(n)} - \hat{y}^{(n)})) \end{pmatrix}$$

$$\frac{\partial \mathcal{J}}{\partial \vec{z}_1} \cdot \frac{\partial \vec{z}_1}{\partial \theta} \Rightarrow \mathrm{sum}\left(\left\{(\vec{\alpha} \circ \vec{\beta}) \circ \vec{\gamma}\right\} \circ \frac{\partial \vec{z}_1}{\partial \theta}\right) = \mathrm{Sum}\left(\begin{pmatrix} \frac{1}{n} \cdot (-2(y^{(1)} - \hat{y}^{(1)})) \\ \frac{1}{n} \cdot (-2(y^{(2)} - \hat{y}^{(2)})) \\ \vdots \\ \frac{1}{n} \cdot (-2(y^{(n)} - \hat{y}^{(n)})) \end{pmatrix} \circ \begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{pmatrix}\right)$$

$$= \mathrm{sum}\left(\begin{pmatrix} \frac{1}{n} \cdot (-2x^{(1)}(y^{(1)} - \hat{y}^{(1)})) \\ \frac{1}{n} \cdot (-2x^{(2)}(y^{(2)} - \hat{y}^{(2)})) \\ \vdots \\ \frac{1}{n} \cdot (-2x^{(n)}(y^{(n)} - \hat{y}^{(n)})) \end{pmatrix}\right) = \frac{1}{n}\sum_{i=1}^{n}\left[-2x^{(i)}(y^{(i)} - \hat{y}^{(i)})\right]$$

따라서 1), 2)의 결과를 이용하여 Question.3-16의 과정을 Hadamard product로 치환할 수 있다.