<Mathematics for Artificial Intelligence> September 28, 2021
SW, Hallym, Fall 2021 Due: 3:00PM, October 5

Problem Set #1

경고: 숙제는 마감일 이후에는 채점하지 않습니다. 답을 작성할 때에는, 본인이 생각한 과정과 이유를 가장 자세하게 작성하세요. 제출 할때 파일명 형식은 "파일명_학번_이름.pdf"로 제출하세요.

이름: 황명원
학번: 20185309

1. (Transpose) Matrix $A \in \mathbb{R}^{m\times n}$이

$$A = \begin{bmatrix} a^{(1)} & \cdots & a^{(n)} \end{bmatrix}$$

와 같이 주어졌다고 가정합니다.

$$A^T = \begin{bmatrix} (a^{(1)})^T \\ \vdots \\ (a^{(n)})^T \end{bmatrix}$$

임을 보이세요.

Sol) 행렬A는 row가 m, column 이 n인 행렬이므로

$$A = \begin{bmatrix} a_1^{(1)} & a_1^{(2)} & \cdots & a_1^{(n)} \\ a_2^{(1)} & a_2^{(2)} & \cdots & a_2^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ a_m^{(1)} & a_m^{(2)} & \cdots & a_m^{(n)} \end{bmatrix}$$

과 같이 나타낼수 있습니다.

A^T는 Transpose 에 정의에 의해 $A^T \in \mathbb{R}^{n\times m}$이 되므로

$$\begin{aligned} A^T &= \begin{bmatrix} a_1^{(1)} & a_2^{(1)} & \cdots & a_m^{(1)} \\ a_1^{(2)} & a_2^{(2)} & \cdots & a_m^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ a_1^{(n)} & a_2^{(n)} & \cdots & a_m^{(n)} \end{bmatrix} \\ &= \begin{bmatrix} (a^{(1)})^T \\ \vdots \\ (a^{(n)})^T \end{bmatrix} \end{aligned}$$

즉 위 가정에서 A^T는 성립합니다.

2. (Matrix Multiplication) Matrix $A \in \mathbb{R}^{3\times 2}$와 $B \in \mathbb{R}^{3\times 2}$는

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

와 같이 주어졌다고 가정합니다. $B^T A$를 구하세요.
Sol)

$$\begin{aligned} B^T A &= \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 1*1+3*1+5*1 & 2*1+4*1+6*1 \\ 1*2+3*1+5*1 & 2*2+4*1+6*1 \end{bmatrix} \\ &= \begin{bmatrix} 9 & 12 \\ 10 & 14 \end{bmatrix} \end{aligned}$$

3. (Matrix 차원 분석) 다음 문제에서 matrix의 차원을 분석하세요.

(a) $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \end{bmatrix}$ 일때 $A \in \mathbb{R}^{\square \times \square}$?

(b) $B \in \mathbb{R}^{4\times 3}$, $C \in \mathbb{R}^{3\times 8}$ 일때, $BC \in \mathbb{R}^{\square \times \square}$?

(c) $s_i \in \mathbb{R}$, $D_i \in \mathbb{R}^{3\times 4}$, $i = 1, \ldots, n$, 임을 가정한다면, $\sum_{i=1}^n s_i D_i \in \mathbb{R}^{\square \times \square}$?

Sol)

(a) A는 row가 2, column 이 3인 행렬이므로 $A \in \mathbb{R}^{2\times 3}$

(b) BC는 행렬의 곱 정의에 의해 $BC \in \mathbb{R}^{4\times 8}$

(c) s_i는 실수 집합이고 D_i는 row가 3 column이 4인 행렬이다.

$\sum_{i=1}^n s_i D_i$는 실수와 행렬의 곱을 i는 1부터 n까지의 합을 구한것이므로

$\sum_{i=1}^n s_i D_i \in \mathbb{R}^{3\times 4}$

4. (Matrix elements) 다음과 같이 정의된 matrix, vector에 대해서 문제를 풀어보세요.

$$a^{(i)} \in \mathbb{R}^n = \left[a_1^{(i)}, \ldots, a_n^{(i)}\right]^T, \quad i = 1, \ldots, m$$

$$A = \begin{bmatrix} (a^{(1)})^T \\ (a^{(2)})^T \\ \vdots \\ (a^{(m)})^T \end{bmatrix}$$

(a) $A \in \mathbb{R}^{\square \times \square}$?

(b) $A_{23} =$?

(c) $A_{4,:} =$?

(d) $A_{:,1}=?$

Sol)

(a) $(a^{(i)})^T = \left[a_1^{(i)}, \ldots, a_n^{(i)}\right]$ 이므로

A= $\begin{bmatrix} a_1^{(1)} & \ldots & a_n^{(1)} \\ a_1^{(2)} & \ldots & a_n^{(2)} \\ \vdots & \ddots & \vdots \\ a_1^{(m)} & \ldots & a_n^{(m)} \end{bmatrix}$ (1)

즉 row는 m, column은 n 인 행렬 형태 이므로

$A \in \mathbb{R}^{m \times n}$

(b) $A_{23} = a_3^{(2)}$ ((1)에 의해)

(c) $A_{4,:} = \left[a_1^{(4)} \quad \ldots \quad a_n^{(4)}\right]$ ((1)에 의해)

(d) $A_{:,1} = \begin{bmatrix} a_1^{(1)} \\ a_1^{(2)} \\ \vdots \\ a_1^{(m)} \end{bmatrix}$ ((1)에 의해)

5. (Matrix Product) $B \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^n$ 이고,

$$\begin{aligned} B &= [b^{(1)}, \ldots, b^{(n)}] \\ b^{(i)} &= [b_1^{(i)}, \ldots, b_m^{(i)}]^T, \quad i = 1, \ldots, n \\ y &= [y_1, \ldots, y_n]^T \end{aligned}$$

이라고 가정할때, 다음 관계를 증명하세요.

$$By = b^{(1)}y_1 + b^{(2)}y_2 + \cdots + b^{(n)}y_n$$

(hint: 정의로부터 출발하여 분배법칙 사용 i.e., $(bc + ec) = c(b + e)$)

Sol) 위에 가정에서 B와 b^i의 관계가 나와있으므로

$$B = \begin{bmatrix} b_1^{(1)} & b_1^{(2)} & \cdots & b_1^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ b_m^{(1)} & b_m^{(2)} & \cdots & b_m^{(n)} \end{bmatrix}$$

따라서 By를 풀어보면(분배법칙 사용)

$$\begin{aligned} By &= \begin{bmatrix} b_1^{(1)} & b_1^{(2)} & \cdots & b_1^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ b_m^{(1)} & b_m^{(2)} & \cdots & b_m^{(n)} \end{bmatrix} \times y \\ &= \begin{bmatrix} b_1^{(1)} \\ \vdots \\ b_m^{(1)} \end{bmatrix} y_1 + \begin{bmatrix} b_1^{(2)} \\ \vdots \\ b_m^{(2)} \end{bmatrix} y_2 + \cdots + \begin{bmatrix} b_1^{(n)} \\ \vdots \\ b_m^{(n)} \end{bmatrix} y_n \end{aligned}$$

이때 가정에서

$$b^{(i)} = [b_1^{(i)}, \ldots, b_m^{(i)}]^T, \quad i = 1, \ldots, n$$

이라고 했으므로

$$By = b^{(1)} y_1 + b^{(2)} y_2 + \cdots + b^{(n)} y_n$$

즉 위 관계는 성립한다.

6. (Machine Learning Model) 다음과 같이 m개의 sample vector가 존재한다고 가정합니다.

$$s^{(i)} = [s_1^{(i)}, \ldots, s_n^{(i)}]^T \in \mathbb{R}^n, \quad i = 1, \ldots, m$$

주어진 sample vector를 다음 함수의 입력으로 사용하려고 합니다. 우선 함수의 정의는 다음과 같습니다:

$$h_\theta(x) = \sum_{i=1}^{n} \theta_i x_i + \theta_0 \tag{2}$$

여기서 함수의 정의역 $\mathbb{R}^n$이며, 공역은 $\mathbb{R}$ 입니다. 위 함수에서 입력에 곱해지는 계수의 (θ)는 parameter 라고 부르며,

$$\theta \in \mathbb{R}^{n+1} = [\theta_0, \theta_1, \ldots, \theta_n]^T$$

라고 가정합니다. 본 문제의 취지는 궁극적으로

$$h_\theta(s^{(i)}), \quad i = 1, \ldots, m,$$

즉, m개의 함수 출력 값을 matrix 또는 vector 연산만으로 수행하고자 합니다. 조금 더 구체적으로는

$$\begin{bmatrix} h_\theta(s^{(1)}) \\ h_\theta(s^{(2)}) \\ \vdots \\ h_\theta(s^{(m)}) \end{bmatrix}$$

와 같이 (효율적으로) 저장하고자 합니다.

(a) **(One-trick)** 첫 단계로, $x^{(i)} \in \mathbb{R}^{n+1}$를 다음과 같이 정의합니다:

$$x^{(i)} = \begin{bmatrix} 1 \\ s_1^{(i)} \\ s_2^{(i)} \\ \vdots \\ s_n^{(i)} \end{bmatrix}$$

위 $x^{(i)}$의 정의를 활용하여

$$h_\theta(s^{(i)}) = \theta^T x^{(i)}$$

임을 보이세요.

(b) 두 번째 단계로 다음과 같이 $X \in \mathbb{R}^{m\times(n+1)}$ matrix를 정의합니다:

$$X = \begin{bmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(m)})^T \end{bmatrix} = \begin{bmatrix} 1 & (s^{(1)})^T \\ 1 & (s^{(2)})^T \\ \vdots & \vdots \\ 1 & (s^{(m)})^T \end{bmatrix}$$

즉, 글로 풀어서 얘기하면 sample vector ($s^{(i)}$) 별로 '1'을 덧댄 벡터 ($x^{(i)}$)의 전치를 row로 구성한 matrix입니다. 위 정의를 사용하여

$$\begin{bmatrix} h_\theta(s^{(1)}) \\ h_\theta(s^{(2)}) \\ \vdots \\ h_\theta(s^{(m)}) \end{bmatrix} = X\theta$$

임을 보이세요.

Sol)

(a) 먼저 $h_\theta(s^{(i)})$를 위 정의대로 풀어보면

$$\begin{aligned} h_\theta(s^{(i)}) &= \sum_{i=1}^{n} \theta_i s_i^{(i)} + \theta_0 \\ &= \theta_0 + \theta_1 s_1^{(i)} + \cdots + \theta_n s_n^{(i)} \end{aligned}$$

그다음 $\theta^T x^{(i)}$ 를 위 정의대로 풀어보면

$$\begin{aligned} \theta^T x^{(i)} &= \begin{bmatrix} \theta_0 & \theta_1 & \cdots & \theta_n \end{bmatrix} \times \begin{bmatrix} 1 \\ s_1^{(i)} \\ \vdots \\ s_n^{(i)} \end{bmatrix} \\ &= \theta_0 + \theta_1 s_1^{(i)} + \cdots + \theta_n s_n^{(i)} \end{aligned}$$

즉 $h_\theta(s^{(i)}) = \theta^T x^{(i)}$ 이므로 위 식은 성립한다.

(b) 우선 X의 정의가 위에 나와있으므로 $X\theta$를 풀어보면

$$
\begin{aligned}
X\theta = \begin{bmatrix} 1 & (s^{(1)})^T \\ 1 & (s^{(2)})^T \\ \vdots & \vdots \\ 1 & (s^{(m)})^T \end{bmatrix} \theta &= \begin{bmatrix} 1 & s_1^{(1)} & \cdots & s_n^{(1)} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & s_1^{(m)} & \cdots & s_n^{(m)} \end{bmatrix} \theta \\
&= \begin{bmatrix} 1 & s_1^{(1)} & \cdots & s_n^{(1)} \\ \vdots & \ddots & \ddots & \vdots \\ 1 & s_1^{(m)} & \cdots & s_n^{(m)} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \\
&= \begin{bmatrix} \theta_0 + \theta_1 s_1^{(1)} + \cdots + \theta_n s_n^{(1)} \\ \vdots \\ \theta_0 + \theta_1 s_1^{(m)} + \cdots + \theta_n s_n^{(m)} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^n \theta_i s_i^{(1)} + \theta_0 \\ \vdots \\ \sum_{i=1}^n \theta_i s_i^{(m)} + \theta_0 \end{bmatrix}
\end{aligned}
$$

이때 $\sum_{i=1}^n \theta_i s_i^{(i)} + \theta_0 = h_\theta(s^{(i)})$ 이므로

$$
\begin{bmatrix} \sum_{i=1}^n \theta_i s_i^{(1)} + \theta_0 \\ \vdots \\ \sum_{i=1}^n \theta_i s_i^{(m)} + \theta_0 \end{bmatrix} = \begin{bmatrix} h_\theta(s^{(1)}) \\ h_\theta(s^{(2)}) \\ \vdots \\ h_\theta(s^{(m)}) \end{bmatrix} = X\theta
$$

즉 위 식은 성립한다.