



- 본 과제는 학회 정규 세션 「Linear Algebra」, 「Basic Statistics」 및 「Math for ML」 일부 내용을 다루며, 개념의 이해와 실제 활용 사례에 대한 이해를 돕기 위해 기획되었습니다. 평가를 위한 것이 아니므로, 주어진 힌트를 적극 활용하시고 학회원 간 토론, Slack의 질의응답을 활용하여 해결해주시요. 단, 답안 표절은 금지합니다.
- 2/11 (수) 23시 59분까지 Github 에 PDF 파일 하나로 제출해주시요. Github에 제출하는 방법을 모른다면 학술부장 혹은 과제 질의응답을 위한 오픈채팅방을 활용해주시요.

## 문제 1 선형대수 기초 (Linear Algebra Basic)

### 1. $Ax = b$ 의 동치 조건

$n \times n$  정사각행렬  $A$ 에 대하여 다음 조건들은 동치(equivalent)입니다.

- (a) 행렬  $A$ 는 가역적(invertible)이다.
- (b) 임의의  $n \times 1$  벡터  $b$ 에 대하여 방정식  $Ax = b$ 는 유일한 해를 갖는다.
- (c) 동차 방정식(Homogeneous equation)  $Ax = 0$ 은 오직 자명한 해(trivial solution,  $x = 0$ )만을 갖는다.

#### 1-1. 다음과 같은 삼변수 함수 $f(x, y, z)$ 가 주어져 있습니다.

$$f(x, y, z) = x^2 + y^2 + 1.5z^2 + xy + yz + zx$$

- (1) 이 함수의 헤시안 행렬(Hessian Matrix)  $A$ 를 구하고 PD(Positive Definite), PSD, ND, NSD 중 무엇인지 판별하시요.
- (2) 위에서 구한 행렬  $A$ 의 역행렬  $A^{-1}$ 을 계산하시요.
- (3) 만약 특정 상수  $k$ 에 대해 함수의 항  $xy$ 가  $kxy$ 로 변하여  $\det(A) = 0$ 이 된다면, 조건 (b)는 여전히 성립하는가? 그 이유를 행렬식(determinant)과 관련지어 설명하시요.

#### 1-2. (Optional)

- (1) (a), (b), (c) 세 조건들이 동치임을 증명하시요.
- (2) 행렬  $A$ 가 가역적이라고 가정할 때, 임의의  $b$ 에 대하여  $Ax = b$ 를 만족하는 해  $x$ 가 항상 존재함을 대수적으로 유도하여 증명하고 해당 증명의 의미를 서술하시요.

1-1.  $f(x,y,z) = x^2 + y^2 + (3z^2 - xy + yz + zx)$

$$H = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

PD, PSD... 는 임의의  $n$ 차원 실수 벡터  $u \neq 0 (u \in \mathbb{R}^n)$  에 대하여

$u^T H u \geq 0 \rightarrow$  부등식에 따라 달라지는 것임. 모든 고유값이 0 초과거나, 미양이나 등과 판단이 있다.

$$\begin{aligned} |H - \lambda I| = 0 &\Rightarrow \begin{vmatrix} 2-\lambda & -1 & 1 \\ -1 & 2-\lambda & 1 \\ 1 & 1 & 3-\lambda \end{vmatrix} = 0 \\ &\Rightarrow (2-\lambda)((2-\lambda)(3-\lambda) - 1) - 1 \cdot (3-\lambda) + 1 \cdot (1 - 2 + \lambda) \\ &= 2-\lambda((2-\lambda)(3-\lambda) - 2) + \lambda - 1 \\ &= -(\lambda-1)(\lambda^2 - 6\lambda + 7) \end{aligned}$$

$$\Rightarrow \lambda_1, \lambda_2, \lambda_3 = 1, 3 \pm \sqrt{2} > 0 \Rightarrow H \text{ is PD}$$

1x (실베스터 판정법)

$$H = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix} \quad \begin{array}{l} \textcircled{1} > 0? \text{ YES} \\ \det(\textcircled{2}) > 0? \text{ YES} \\ \det(\textcircled{3}) > 0? \text{ YES} \end{array} \Rightarrow \text{PD}$$

(i) 모든 대칭행렬에 대해  $LDL^T$  분해 가능. ( $L$ : 하삼각 행렬,  $D$ : 대각행렬)

(ii)  $x^T A x = x^T L D L^T x = (Lx)^T D (Lx) = y^T D y = \sum_{i=1}^n d_i y_i^2$ ,  $y_i^2$ 은 항상 양수므로,  $d_i$ 가 중요!  
 PD 판단대상.  $\downarrow$   
 let  $Lx = y$

(iii)  $\det(A_k) = d_1 \times d_2 \times \dots \times d_k$  ( $\det(A_k)$ 은  $k \times k$ 번째 주대행렬)

$$(2) H^{-1} = \frac{1}{\det(H)} \cdot C = \begin{bmatrix} 5/7 & -2/7 & -1/7 \\ -2/7 & 5/7 & 1/7 \\ -1/7 & 1/7 & 3/7 \end{bmatrix}$$

$$\det(H) = 2 \cdot (6-1) - 1 \cdot (3-1) + 1 \cdot (1-2) = 7$$

$$C = \begin{bmatrix} 5 & -2 & -1 \\ -2 & 5 & 1 \\ -1 & 1 & 3 \end{bmatrix}$$

(3)  $\det(A) = 0 \Leftrightarrow$  행렬  $A$ 는 역행렬을 가질 수 없다 (비가역적)

$\Leftrightarrow$  방정식  $Ax=b$ 는 유일해를 갖는다 (x)  $\because$  해를 갖기위해  $x=A^{-1}b$ 가 가능해야 but  $A^{-1}(x)$

## 문제 2 특이값 분해 (SVD)

특이값 분해(SVD)는 대각화와 달리 모든 크기의 행렬에 대해 적용이 가능하며, 그 계산 과정은 다음과 같습니다.

- 1) 항상 대칭 행렬을 이루는 두 행렬  $B^T B, B B^T$  를 계산하고, 이들의 고유값을 계산해 특이값  $\sigma$ 와 고유 벡터를 통해 직교대각화하여  $B^T B = V D V^T, B B^T = U D U^T$  를 구합니다.
- 2) 0이 아닌 특이값들을 내림차순으로 나열하여  $\Sigma$ 를 구성하고, 이들을 모두 활용해  $A = U \Sigma V^T$  를 구합니다.

다음 행렬  $B$ 에 특이값 분해를 적용하였을 때 나오는 행렬  $U, \Sigma, V^T$  를 각각 구하시오.

$$B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

① 고유값 ② 특이값 ③ 고유벡터  $(A - \lambda I)x = 0$

$$B B^T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \lambda = 2, 1 \quad \sigma_1 = \sqrt{2}, \sigma_2 = 1$$

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$v_i = \frac{1}{\sigma_i} B^T u_i \quad (\text{SVD의 경우 } B = U \Sigma V^T)$$

$$v_1 = \frac{1}{\sigma_1} B^T u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$v_2 = \frac{1}{\sigma_2} B^T u_2 = \frac{1}{1} \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$V^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

$$v_3 = v_1 \times v_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

$$\therefore U \Sigma V^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

### 문제 3 Convex Sets & Functions

#### 3-1. Convex Set

1. 다음 집합들이 Convex set 인지 증명하시오.

$$C_1 = \{(x, y) \in \mathbb{R}^2 : x + y \leq 1\}$$

$$C_2 = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$$

$$C_3 = \{(x, y) \in \mathbb{R}^2 : y \geq e^x\}$$

2. 함수  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  가 convex function 일 때,  $f$  의 epigraph 인 집합  $S$  가 convex set 임을 보이시오.

$$S = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : t \geq f(x)\}$$

#### 3-2. Convex Function

1. Convex 함수의 합성

$f : \mathbb{R}^d \rightarrow \mathbb{R}$  가 실수 값을 갖는 함수라고 합시다.

(a) 함수  $f(x)$  와  $g(x)$  가 convex 일 때,  $f(x) + g(x)$  또한 convex 임을 보이시오.

(b)  $A$  와  $b$  가 호환되는 크기의 행렬과 벡터라고 합시다.  $f(x)$  가 convex 라면  $f(Ax + b)$  또한 convex 임을 보이시오.

2. Convex Optimization

또 다른 함수  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  를 고려해 봅시다.

$$f(x) = \|Ax - b\|^2.$$

오직 convex 함수의 정의만을 사용하여,  $f(x)$  가  $x \in \mathbb{R}^d$  에서 convex 임을 보이시오.

$$\begin{aligned} u &= Ax - b & \text{then } g(\theta u + (1-\theta)v) &\leq \theta g(u) + (1-\theta)g(v) \\ v &= Ay - b \\ g(x) &= \|x\|^2 & \theta \|u\|^2 + (1-\theta)\|v\|^2 - \|\theta u + (1-\theta)v\|^2 \\ & &= \theta \|u\|^2 + (1-\theta)\|v\|^2 - [\theta^2 \|u\|^2 + (1-\theta)^2 \|v\|^2 + 2\theta(1-\theta)u'v] \\ & &= \theta(1-\theta)(\|u\|^2 + \|v\|^2 - 2u'v) \\ & &= \theta(1-\theta)\|u-v\|^2 \geq 0 \\ & &\therefore f(x) \text{ is convex func} \end{aligned}$$

$$C_1 = \{ (x, y) \in \mathbb{R}^2 : x+y \leq 1 \}$$

$$(x_1, y_1), (x_2, y_2) \in C_1 \Rightarrow x_1+y_1 \leq 1, x_2+y_2 \leq 1$$

$$\text{If } (x_{\text{new}}, y_{\text{new}}) = \underbrace{\theta(x_1+y_1)}_{\leq 1} + (1-\theta) \underbrace{(x_2+y_2)}_{\leq 1} \in C_1 \Rightarrow \text{Convex Set}$$

$$\therefore \theta(x_1+y_1) + (1-\theta)(x_2+y_2) \leq \theta \cdot 1 + (1-\theta) \cdot 1 = 1$$

$$C_2 = \{ x \in \mathbb{R}^n : \|x\|_1 \leq 1 \}$$

$$x, y \in C_2 \Rightarrow \|x\|_1 \leq 1, \|y\|_1 \leq 1$$

$$\|\theta x + (1-\theta)y\|_1 \leq \theta \|x\|_1 + (1-\theta) \|y\|_1 \leq 1$$

$$\hookrightarrow \text{삼각 부등식: } \|a+b\| \leq \|a\| + \|b\|$$

$$C_3 = \{ (x, y) \in \mathbb{R}^2 : y \geq e^x \}$$

$$(x_1, y_1), (x_2, y_2) \in C_3 \Rightarrow y_1 \geq e^{x_1}, y_2 \geq e^{x_2}$$

$$\theta y_1 + (1-\theta)y_2 \geq \theta e^{x_1} + (1-\theta)e^{x_2} \geq e^{\theta x_1 + (1-\theta)x_2}$$

$$(x_1, t_1), (x_2, t_2) \in S \Rightarrow t_1 \geq f(x_1), t_2 \geq f(x_2)$$

$$x' = \theta x_1 + (1-\theta)x_2, t' = \theta t_1 + (1-\theta)t_2$$

$\hookrightarrow$  Convex func

$$f(x') \leq \theta f(x_1) + (1-\theta)f(x_2)$$

$$\leq \theta t_1 + (1-\theta)t_2 = t'$$

$$\Rightarrow t' \geq f(x') \quad \text{바깥쪽에 S이 포함됨} \therefore S \text{ is Convex Set}$$

$$f, g \text{ is convex} \Rightarrow f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \quad \text{만약}$$

$$f+g \text{ 하의 부등식 유지되므로 Convex.}$$

$$\text{let } h(x) = f(Ax+b), h(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + b) = f(\theta(Ax+b) + (1-\theta)(Ay+b))$$

$$= f(\theta \underbrace{X}_{x} + (1-\theta) \underbrace{Y}_{y}) \leq \theta f(x) + (1-\theta)f(y)$$

$\hookrightarrow f \text{ is convex func}$

$$= \theta h(x) + (1-\theta)h(y)$$

$$\therefore h(x) \text{ is Convex func.}$$

## 문제 4 정보이론 (Information Theory)

### 4-1. Entropy

확률분포  $p(x, y)$  가 다음과 같이 주어졌다고 합시다.

$X \backslash Y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

다음 값들을 구하시오.

(a)  $H(X), H(Y)$ .

(b)  $H(X | Y), H(Y | X)$ .

(c)  $H(X, Y)$ .

(d)  $H(Y) - H(Y | X)$ .

(e)  $I(X; Y)$ .

(f) (a)부터 (e)까지 구한 양(quantity)들을 벤 다이어그램으로 나타내시오.

### 4-2. KL-divergence

(a)  $D(q||p) = D(p||q)$  가 성립하지 않는 반례를 제시하시오.

(b)  $D(p||q) \geq 0$ 임을 보이시오. (Hint : Jensen's Inequality)

---

## Reference

- Elements of Information Theory (2nd Ed.), T. M. Cover & J. A. Thomas
- Convex Optimization for ML, Prof. Changho Suh

## Data Science Lab

담당자: 14 기 박창용, 어희정, 여준호

qkrckddyd0@yonsei.ac.kr

heejung.uh@gmail.com

asap03153@gmail.com

4-1

$X \backslash Y$	0	1
0	$\frac{1}{3}$	$\frac{1}{3}$
1	0	$\frac{1}{3}$

$X = 0$  w.p  $\frac{2}{3}$   
 $\quad \quad 1 \quad \cdot \quad \frac{1}{3}$   
 $Y = 0$  w.p  $\frac{1}{3}$   
 $\quad \quad 1 \quad \cdot \quad \frac{2}{3}$

$$H(X) = - \sum_x p(x) \log p(x) = -\frac{2}{3} (\log 2 - \log 3) - \frac{1}{3} (\log 1 - \log 3) = \log 3 - \frac{2}{3}$$

$$H(Y) = - \sum_y p(y) \log p(y) = \log 3 - \frac{2}{3} \quad \therefore X \text{과 } Y \text{의 분포 동일.}$$

$$H(X|Y) = p(y=0) H(X|y=0) + p(y=1) H(X|y=1) = \frac{2}{3} \cdot 1 = \frac{2}{3}$$

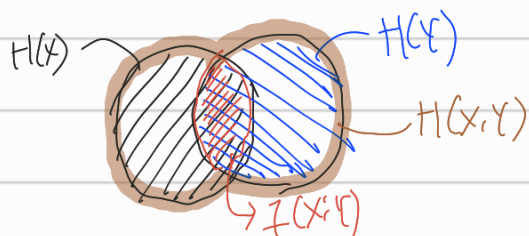
$$\therefore p(X=0|Y=1) = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

$$H(Y|X) = p(X=0) H(Y|X=0) + p(X=1) H(Y|X=1) = \frac{2}{3} \cdot 1 = \frac{2}{3}$$

$$H(X, Y) = H(X) + H(Y|X) = \log 3$$

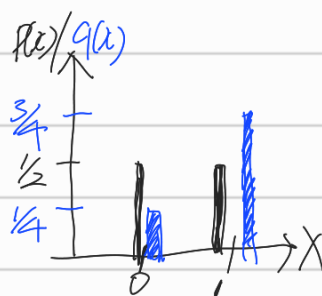
$$H(Y) - H(Y|X) = \log 3 - \frac{2}{3}$$

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = \log 3 - \frac{2}{3}$$



4-2

$$D(P||Q) := \sum_x p(x) \log \frac{p(x)}{q(x)}$$



$$D(P||Q) = \frac{1}{2} \log \frac{1/2}{1/4} + \frac{1}{2} \log \frac{1/2}{3/4} = \frac{1}{2} + \frac{1}{2} - \frac{1}{2} \log 3$$

$$D(Q||P) = \frac{1}{4} \log \frac{1/4}{1/2} + \frac{3}{4} \log \frac{3/4}{1/2} = -\frac{1}{4} + \frac{3}{4} \log 3 - \frac{3}{4} \quad (\neq)$$

$$D(P||Q) \geq 0$$

$$D(P||Q) := \sum_x p(x) \log \frac{p(x)}{q(x)} = E_P \left[ \log \frac{p(x)}{q(x)} \right] = -E_P \left[ \log \frac{q(x)}{p(x)} \right]$$

$$\rightarrow -\log \left( E_P \left[ \frac{q(x)}{p(x)} \right] \right) = -\log 1 = 0$$

Jensen's

$f(x)$   $\nearrow$  shape

$$E f(x) < f(E x)$$

Jensen's Ineq

$$\rightarrow E_P \left[ \frac{q(x)}{p(x)} \right] = \sum_x p(x) \frac{q(x)}{p(x)} = \sum_x q(x) = 1$$