



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

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.

- 행렬 A 는 가역적(invertible)이다.
- 임의의 $n \times 1$ 벡터 b 에 대하여 방정식 $Ax = b$ 는 유일한 해를 갖는다.
- 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.

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

- 이 함수의 헤시안 행렬(Hessian Matrix) A 를 구하고 PD(Positive Definite), PSD, ND, NSD 중 무엇인지 판별하시오.
- 위에서 구한 행렬 A 의 역행렬 A^{-1} 을 계산하시오.
- 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 설명하시오.

1-2. (Optional)

- (a), (b), (c) 세 조건들이 동치임을 증명하시오.
- 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 서술하시오.

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

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

PD, PSD, ... 는 임의의 n 차원 실수 빼어 $H \neq 0$ ($\lambda \in \mathbb{R}^n$) 이 아하여

$\lambda_1 H \lambda_1 \geq 0$ \rightarrow 뿐만 아니라 달라지는 것이다. 모든 고유값이 0 초과이고, 이중이거나 등차 일관성이 있다.

$$\begin{aligned} |H - \lambda I| = 0 \Rightarrow \begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 3-\lambda \end{vmatrix} = 0 &\Rightarrow (2-\lambda)[(2-\lambda)(3-\lambda)-1] - 1 \cdot (3-\lambda) + 1 \cdot (1-2+\lambda) \\ &= 2\lambda[(2-\lambda)(3-\lambda)-2] + \lambda - 1 \\ &= -(\lambda-1)(\lambda^2 - 6\lambda + 7) \\ \Rightarrow \lambda_1, \lambda_2, \lambda_3 &= 1, 3 \pm \sqrt{2} > 0 \Rightarrow H \text{ is PD} \end{aligned}$$

+x (실버스터 단점법)

$$H = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

① $\det(\underline{\underline{2}}) > 0$? Yes
 ② $\det(\underline{\underline{1}}) > 0$? Yes \Rightarrow PD
 ③ $\det(\underline{\underline{3}}) > 0$? Yes

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

(ii) $\underline{\underline{x}}' \underline{\underline{A}} \underline{\underline{x}} = \underline{\underline{x}}' \underline{\underline{L}} \underline{\underline{D}} \underline{\underline{L}}' \underline{\underline{x}} = (\underline{\underline{L}}' \underline{\underline{x}})' \underline{\underline{D}} (\underline{\underline{L}}' \underline{\underline{x}}) = \underline{\underline{y}}' \underline{\underline{D}} \underline{\underline{y}} = \sum_{i=1}^n d_i y_i^2$, y_i 는 항상 양수므로, d_i 가 중요!
 PD 판단 단계.

(iii) $\det(A_k) = d_1 \times d_2 \times \dots \times d_k$ ($\det(A_k)$ 는 $k \times k$ 빼제 주요행렬식)

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

$$\det(H) = 2 \cdot (6) - 1 \cdot (3) + 1 \cdot (1-2) = 7 \quad C = \begin{bmatrix} 5 & -2 & -1 \\ -2 & 5 & -1 \\ -1 & 1 & 3 \end{bmatrix}$$

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

\Leftrightarrow 방정식 $Ax = b$ 는 유일한 해를 갖는다 (x) \because 해를 찾기 위해 $x = A^{-1}b$ 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 $B^T B, BB^T$ 를 계산하고, 이들의 고유값을 계산해 특이값 σ 와 고유 벡터를 통해 직교대각화하여 $B^T B = V D V^T, BB^T = U D U^T$ 를 구합니다.
- 2) 0이 아닌 특이값들을 내림차순으로 나열하여 Σ 를 구성하고, 이를 모두 활용해 $A = U \Sigma V^T$ 를 구합니다.

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

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

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

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

$$U_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad U_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$V_i = \frac{1}{\sigma_i} B' u_i \quad (\text{SVD의 정의 } B = U \Sigma V')$$

$$V_1 = \frac{1}{\sqrt{2}} B' u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$V_2 = \frac{1}{\sqrt{2}} B' u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad V' = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

$$V_3 = V_1 \times V_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$\therefore U \Sigma V' = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

문제 3 Convex Sets & Functions

3-1. Convex Set

1. 다음 집합들이 Convex set인지 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 function일 때, f 의 epigraph인 집합 S 가 convex set임을 보이시오.

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

3-2. Convex Function

1. Convex 함수의 합성

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

- (a) 함수 $f(x)$ 와 $g(x)$ 가 convex일 때, $f(x) + g(x)$ 또한 convex임을 보이시오.

- (b) A 와 b 가 호환되는 크기의 행렬과 벡터라고 합시다. $f(x)$ 가 convex라면 $f(Ax + b)$ 또한 convex임을 보이시오.

2. Convex Optimization

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

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

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

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

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

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

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

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

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

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

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

\hookrightarrow 삼각부등식: $\|a+b\| \leq \|a\| + \|b\|$

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

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

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

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

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

\hookrightarrow Convex func

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

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

$\Rightarrow t' \geq f(x) \quad \text{내분할도 } S \text{의 특성} \therefore S \text{ is Convex Set}$

f, g is convex. $\Rightarrow f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y)$ 만족

$f+g$ 하도 부등호 유지되므로 Convex.

$$\begin{aligned} \text{let } h(x) &= f(Ax+b), h(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + b) = f(\theta \underline{(Ax+b)} + (1-\theta) \underline{(Ay+b)}) \\ &= f(\theta X + (1-\theta)Y) \stackrel{f \text{ is convex func}}{\leq} \theta f(X) + (1-\theta)f(Y) \\ &= \theta h(x) + (1-\theta)h(y) \\ \therefore h(x) &\text{ is Convex func.} \end{aligned}$$

문제 4 정보이론 (Information Theory)

4-1. Entropy

확률분포 $p(x, y)$ 가 다음과 같이 주어졌다고 합시다.

$X \setminus Y$	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	$\begin{array}{c cc} X & 0 & 1 \\ \hline 0 & \frac{1}{3} & \frac{1}{3} \\ 1 & \frac{1}{3} & \frac{1}{3} \end{array}$	$X = 0 \text{ w.p } \frac{2}{3}$
		$Y = 0 \text{ w.p } \frac{1}{3}$

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

$$H(Y) = -\sum_y p(y) \log p(y) = \log 3 - \frac{2}{3} \because X \text{과 } Y \text{는 서로 독립적입니다.}$$

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

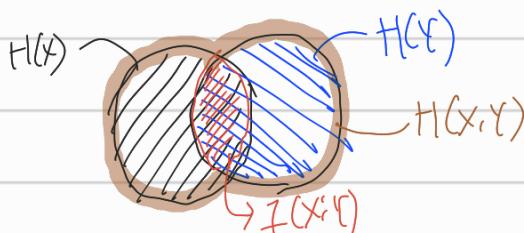
$$\therefore P(X=0|Y=1) = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$$

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

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

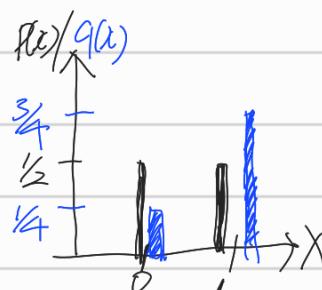
$$H(Y) - H(Y|X) = \log 3 - \frac{2}{3}$$

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



4-2

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



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

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

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

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

Jensen's
 $f(x)$ \nwarrow shape

$$E[f] < f(E)$$

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

Jensen's Ineq

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