기수: 15기　　이름: 최 현

문제 1　선형대수 기초 (Linear Algebra Basic)

1. $Ax = b$의 동치 조건

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

(a) 행렬 A는 가역적(invertible)이다.

(b) 임의의 $n \times 1$ 벡터 b에 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) A를 구하고 PD(Positive Definite), PSD, ND, NSD 중 무엇인지 판별하시오.

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

(3) 만약 특정 상수 k에 대해 함수의 항 xy가 kxy로 변하여 $\det(A) = 0$이 된다면, 조건 (b)는 여전히 성립하는가? 그 이유를 행렬식(determinant)과 관련지어 설명하시오.

(1)
$$A = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 z}{\partial z^2} \end{pmatrix} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

Let $x := \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. Then $x^T A x = (x_1 \ x_2 \ x_3) \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (x_1 \ x_2 \ x_3) \begin{pmatrix} 2x_1 + x_2 + x_3 \\ x_1 + 2x_2 + x_3 \\ x_1 + x_2 + 3x_3 \end{pmatrix}$

$$= x_1(2x_1 + x_2 + x_3) + x_2(x_1 + 2x_2 + x_3) + x_3(x_1 + x_2 + 3x_3)$$

$$= 2x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 2x_2x_3 + 2x_3x_1$$

$$= (x_1 + x_2 + x_3)^2 + x_1^2 + x_2^2 + 2x_3^2 > 0 \quad (\because x \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix})$$

$\Rightarrow$ Thus A is PD

(2) $\det(A) = 2(6-1) - 1(3-1) + 1(1-2) = 10 - 2 - 1 = 7$

Cofactor Matrix $C = \begin{pmatrix} 5 & -2 & -1 \\ -2 & 5 & -1 \\ -1 & -1 & 3 \end{pmatrix} \Rightarrow C^T = \begin{pmatrix} 5 & -2 & -1 \\ -2 & 5 & -1 \\ -1 & -1 & 3 \end{pmatrix}$

$$\therefore A^{-1} = \frac{1}{\det(A)} C^T = \begin{pmatrix} \frac{5}{7} & -\frac{2}{7} & -\frac{1}{7} \\ -\frac{2}{7} & \frac{5}{7} & -\frac{1}{7} \\ -\frac{1}{7} & -\frac{1}{7} & \frac{3}{7} \end{pmatrix}$$

(3) Suppose $\det(A)=0$.
$\Updownarrow$
Then A is not invertible since $A^{-1}=\frac{1}{\det(A)}C^T$.
$\Updownarrow$
(b)

1-2. (Optional)

(1) (a), (b), (c) 세 조건들이 동치임을 증명하시오.

(2) 행렬 A가 가역적이라고 가정할 때, 임의의 b에 대하여 $Ax = b$를 만족하는 해 x가 항상 존재함을 대수적으로 유도하여 증명하고 해당 증명의 의미를 서술하시오.

(1) pf (a)⇒(b)

Suppose A is invertible.
Then A^{-1} exists. Suppose $Ax=b$ where $b=\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$. Then $Ax=b \Leftrightarrow A^{-1}(Ax)=A^{-1}(b) \Leftrightarrow Ix=A^{-1}b \Leftrightarrow x=A^{-1}b$

Claim: x is unique.

pf Suppose $\exists x_1, x_2$ s.t. $Ax_1=b$ and $Ax_2=b$. This implies $A(x_1-x_2)=\underset{\text{zero vector}}{0}$ and multiplying A^{-1} both sides yields $x_1-x_2=A^{-1}0=0 \Leftrightarrow x_1=x_2$. Thus x is unique.

pf (b)⇒(c)

Suppose $Ax=b$ has unique vector x where b is any $n\times 1$ vector.
Fix $b=0$ ($n\times 1$ zero vector). Then done.

pf (c) ⇒ (a)

Suppose (c). Then A is injective since null space of $A=\{0\}$. $\Leftrightarrow$ A is surjective $\Rightarrow \forall b, \exists x$ s.t. $Ax=b$
$\Rightarrow A$ is invertible.

(2) By assumption A is invertible. So we can define $x:=A^{-1}b$ for any given $n\times 1$ vector b.
Then $Ax=A(A^{-1}b)=Ib=b$. Thus the solution always exists.

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^TB, BB^T를 계산하고, 이들의 고유값을 계산해 특이값 σ와 고유벡터를 통해 직교대각화하여 $B^TB = VDV^T, BB^T = UDU^T$를 구합니다.
>
> 2) 0이 아닌 특이값들을 내림차순으로 나열하여 Σ를 구성하고, 이들을 모두 활용해 $A = U\Sigma V^T$를 구합니다.

다음 행렬 B에 특이값 분해를 적용하였을 때 나오는 행렬 U, Σ, V^T를 각각 구하시오.

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

$BB^T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ ⇒ 상삼각행렬의 대각성분 = eigenvalue ⇒ $\lambda = 1, 2$ ⇒ $\sigma = 1, \sqrt{2}$ ⇒ $\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

$BB^T v = \lambda v \rightarrow$ (1) $(\lambda=1)$ $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Leftrightarrow \begin{matrix} 2v_1 = v_1 \\ v_2 = v_2 \end{matrix}$ ⇒ 대응하는 단위 고유벡터 $= \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

(2) $(\lambda=2)$ $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2v_1 \\ 2v_2 \end{pmatrix} \Leftrightarrow \begin{matrix} 2v_1 = 2v_1 \\ v_2 = 2v_2 \end{matrix}$ ⇒ 대응하는 단위 고유벡터 $= \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

⇒ $U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$B^TB = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ ⇒ $\det(B^TB - \lambda I) = \det\left(\begin{pmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix}\right) = (1-\lambda)^3 - (1-\lambda) = (1-\lambda)((1-\lambda)^2 - 1) = 0$

$\Downarrow$

$\lambda = 0, 1, 2$

$B^TB = \lambda v \rightarrow$ (1) $(\lambda=0)$ $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{matrix} v_1 + v_2 = 0 \\ v_1 + v_2 = 0 \\ v_3 = 0 \end{matrix}$ ⇒ 대응하는 단위 고유벡터 $= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

(2) $(\lambda=1)$ $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \Leftrightarrow \begin{matrix} v_1 + v_2 = v_1 \\ v_1 + v_2 = v_2 \\ v_3 = v_3 \end{matrix}$ ⇒ 대응하는 단위 고유벡터 $= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

(3) $(\lambda=2)$ $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 2v_1 \\ 2v_2 \\ 2v_3 \end{pmatrix} \Leftrightarrow \begin{matrix} v_1 + v_2 = 2v_1 \\ v_1 + v_2 = 2v_2 \\ v_3 = 2v_3 \end{matrix}$ ⇒ 대응하는 단위 고유벡터 $= \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

⇒ $V = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}$

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

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, f의 epigraph인 집합 S가 convex set임을 보이시오.

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

1. Suppose $x, y \in C_1$ and choose any $\lambda \in [0,1]$ where $x=(x_1, x_2), y=(y_1, y_2)$. $\Rightarrow x_1+x_2 \le 1$, $y_1+y_2 \le 1$

Then $w=\lambda x+(1-\lambda)y = (\lambda x_1+(1-\lambda)y_1, \lambda x_2+(1-\lambda)y_2) \in C_1$ since $\lambda x_1+(1-\lambda)y_1+\lambda x_2+(1-\lambda)y_2 = \lambda(x_1+x_2)+(1-\lambda)(y_1+y_2) \le \lambda+1-\lambda=1$

Suppose $x, y \in C_2$. $\Rightarrow \|x\|_1 \le 1$, $\|y\|_1 \le 1$

Then for any $\lambda \in [0,1]$, $\lambda x+(1-\lambda)y \in C_2$ since $\|\lambda x+(1-\lambda)y\|_1 \underset{\text{This holds by triangle inequality.}}{\le} \|\lambda x\|_1+\|(1-\lambda)y\|_1 = \lambda\|x\|_1+(1-\lambda)\|y\|_1 \le \lambda+(1-\lambda)=1$

Suppose $x, y \in C_3$ where $x=(x_1,x_2)$, $y=(y_1,y_2)$. $\Rightarrow x_2 \ge e^{x_1}$, $y_2 \ge e^{y_1}$.

Then for any $\lambda \in [0,1]$, $\lambda x+(1-\lambda)y = (\lambda x_1+(1-\lambda)y_1, \lambda x_2+(1-\lambda)y_2) \in C_3$ since $\lambda x_2+(1-\lambda)y_2 \underset{x,y \in C_3}{\ge} \lambda e^{x_1}+(1-\lambda)e^{y_1} \underset{\text{This holds since } e^x \text{ is convex function.}}{\ge} e^{\lambda x_1+(1-\lambda)y_1}$

---

2. Suppose $p=(p_1,p_2)$, $q=(q_1,q_2)$ and $p,q \in S$. $\Rightarrow p_2 \ge f(p_1)$, $q_2 \ge f(q_1)$

Then for any $\lambda \in [0,1]$, $\lambda p+(1-\lambda)q = (\lambda p_1+(1-\lambda)q_1, \lambda p_2+(1-\lambda)q_2) \in S$ since $\lambda p_2+(1-\lambda)q_2 \underset{p,q\in S}{\ge} \lambda f(p_1)+(1-\lambda)f(q_1) \underset{f \text{ is convex}}{\ge} f(\lambda p_1+(1-\lambda)q_1)$

3-2. Convex Function

1. Convex 함수의 합성

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

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

   (b) A와 b가 호환되는 크기의 행렬과 벡터라고 합시다. $f(x)$가 convex라면 $f(Ax+b)$ 또한 convex 임을 보이시오.

2. Convex Optimization

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

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

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

1.

(a) Suppose $f(x), g(x)$ convex. $\Rightarrow \forall \lambda \in [0,1]$, $x_1, x_2 \in S \Rightarrow f(\lambda x_1+(1-\lambda)x_2) \le \lambda f(x_1)+(1-\lambda)f(x_2)$ where S is convex set.

$g(\lambda x_1+(1-\lambda)x_2) \le \lambda g(x_1)+(1-\lambda)g(x_2)$

Then $f(x)+g(x)$ is convex function since

$f(\lambda x_1+(1-\lambda)x_2)+g(\lambda x_1+(1-\lambda)x_2) \le \lambda f(x_1)+(1-\lambda)f(x_2)+\lambda g(x_1)+(1-\lambda)g(x_2) = \lambda(f(x_1)+g(x_1))+(1-\lambda)(f(x_2)+g(x_2))$

(b) For any $x, y \in \mathbb{R}^d$, $\lambda \in [0,1]$, $f(A(\lambda x+(1-\lambda)y)+b) = f(\lambda Ax+(1-\lambda)Ay+b) = f(\lambda(Ax+b)+(1-\lambda)(Ay+b))$

$\underset{f(x) \text{ is convex}}{\le} \lambda f(Ax+b)+(1-\lambda)f(Ay+b)$

$\therefore f(Ax+b)$ is convex function

2. For any $x, y \in \mathbb{R}^d$, $\lambda \in [0,1]$, $f(\lambda x+(1-\lambda)y) = \|A(\lambda x+(1-\lambda)y)-b\|^2 = \|\lambda(Ax-b)+(1-\lambda)(Ay-b)\|^2$

$\|\cdot\|^2$ is convex $\le \lambda\|Ax-b\|^2+(1-\lambda)\|Ay-b\|^2$

$= \lambda f(x)+(1-\lambda)f(y)$

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) $H(X \mid Y), H(Y \mid X)$.

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

(d) $H(Y) - H(Y \mid X)$.

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

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

(a) $P(X=0) = \frac{2}{3}$, $P(X=1) = \frac{1}{3}$ $\Rightarrow H(X) = -\sum_{x \in X} p(x) \log_2 p(x) = -\frac{2}{3}\log_2 \frac{2}{3} - \frac{1}{3}\log_2 \frac{1}{3}$

$P(Y=0) = \frac{1}{3}$, $P(Y=1) = \frac{2}{3}$ $\Rightarrow H(Y) = -\sum_{y \in Y} p(y) \log_2 p(y) = -\frac{1}{3}\log_2 \frac{1}{3} - \frac{2}{3}\log_2 \frac{2}{3}$

(b)
$$P(X|Y) = \begin{cases} P(0|0) = 1 \\ P(1|0) = 0 \\ P(0|1) = \frac{1}{2} \\ P(1|1) = \frac{1}{2} \end{cases} \Rightarrow H(X|Y) = -\sum_{x,y} p(x,y) \log_2 p(x,y) = -\left(\frac{1}{3}\log_2 \frac{1}{2} + \frac{1}{3}\log_2 \frac{1}{2}\right) = \frac{2}{3}$$

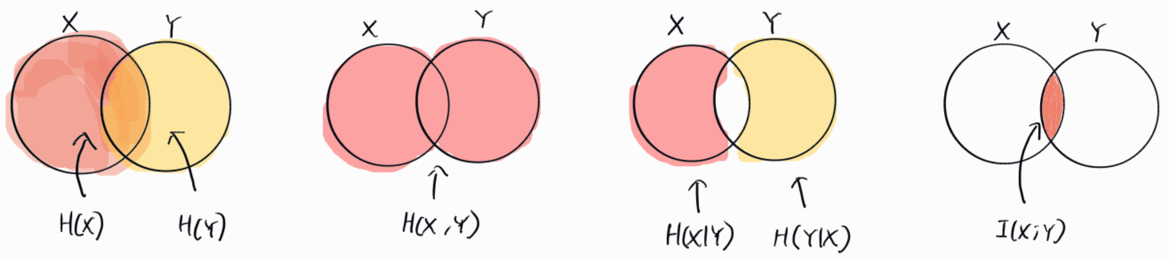
$$P(Y|X) = \begin{cases} P(0|0) = \frac{1}{2} \\ P(1|0) = \frac{1}{2} \\ P(0|1) = 0 \\ P(1|1) = 1 \end{cases} \Rightarrow H(Y|X) = -\sum_{x,y} p(x,y) \log_2 p(x,y) = -\left(\frac{1}{3}\log_2 \frac{1}{2} + \frac{1}{3}\log_2 \frac{1}{2}\right) = \frac{2}{3}$$

(c) $H(X,Y) = H(X) + H(Y|X) = \frac{2}{3} + \frac{1}{3}\log_2 3 + \frac{2}{3}(\log_2 3 - 1) = \log_2 3$

(d) $H(Y) - H(Y|X) = \frac{1}{3}\log_2 3 + \frac{2}{3}(\log_2 3 - 1) - \frac{2}{3} = \log_2 3 - \frac{4}{3}$

(e) $I(X;Y) = H(X) + H(Y) - H(X,Y) = 2(\log_2 3 - \frac{2}{3}) - \log_2 3 = \log_2 3 - \frac{4}{3}$

(f)



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)

(a) In sample space $S=\{0,1\}$, define $p=(p(0),p(1))=(0.9,\ 0.1)$, $q=(q(0),q(1))=(0.5,\ 0.5)$

Then $D(p\|q)=\sum_x p(x)\log\frac{p(x)}{q(x)} = p(0)\log\frac{p(0)}{q(0)} + p(1)\log\frac{p(1)}{q(1)} = 0.9\log\frac{9}{5} + 0.1\log\frac{1}{5}$

$D(q\|p)=\sum_x q(x)\log\frac{q(x)}{p(x)} = q(0)\log\frac{q(0)}{p(0)} + q(1)\log\frac{q(1)}{p(1)} = 0.5\log\frac{5}{9} + 0.5\log 5$

$\Rightarrow D(p\|q) \neq D(q\|p)$

(b) $$D(p\|q) = E_p\left[\log\frac{p(x)}{q(x)}\right] = E_p\left[-\log\frac{q(x)}{p(x)}\right] \geq -\log\left(E_p\left[\frac{q(x)}{p(x)}\right]\right) = \begin{cases} -\log\left(\sum p(x)\cdot\frac{q(x)}{p(x)}\right) = -\log 1 = 0 & (\text{discrete case}) \\ -\log\left(\int p(x)\cdot\frac{q(x)}{p(x)}dx\right) = -\log 1 = 0 & (\text{continuous case}) \end{cases}$$

Since $-\log x$ is convex function, we can use Jensen's inequality

$\therefore D(p\|q) \geq 0$