기수 : 15　　이름 : 현승원

26-1 DSL 정규 세션

# Math for ML

- □ 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 해결해주십시오. 단, 답안 표절은 금지합니다.
- □ 2/11 (수) 23시 59분까지 Github 에 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.

(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 판별하시오.

(2) 위에서 구한 행렬 $A$의 역행렬 $A^{-1}$을 계산하시오.

(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 설명하시오.

1-2. (Optional)

(1) (a), (b), (c) 세 조건들이 동치임을 증명하시오.

(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 서술하시오.

1-1

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

$\det(A-\lambda I) = \begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 3-\lambda \end{vmatrix} = (2-\lambda)\{(2-\lambda)(3-\lambda)-1\} - 1\{(3-\lambda)-1\} + 1\{1-(2-\lambda)\}$

$= (2-\lambda)(5-5\lambda+\lambda^2) - (2-\lambda) + (-1+\lambda)$

$= 10-15\lambda+7\lambda^2-\lambda^3-2+\lambda-1+\lambda = 7-13\lambda+7\lambda^2-\lambda^3 = (1-\lambda)(\lambda^2-6\lambda+7)$

$\lambda = 1,\ 3\pm\sqrt{2}$ 고유값 모두 양수 → PD

(2) $\left[\begin{array}{ccc|ccc} 2&1&1&1&0&0 \\ 1&2&1&0&1&0 \\ 1&1&3&0&0&1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 2&1&1&1&0&0 \\ 2&4&2&0&2&0 \\ 2&2&6&0&0&2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 2&1&1&1&0&0 \\ 0&3&1&-1&2&0 \\ 0&1&5&-1&0&2 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 2&0&-4&2&0&-2 \\ 0&0&-14&2&2&-6 \\ 0&1&5&-1&0&2 \end{array}\right]$

$\to \left[\begin{array}{ccc|ccc} 1&0&-2&1&0&-1 \\ 0&1&5&-1&0&2 \\ 0&0&1&-\frac{1}{7}&-\frac{1}{7}&\frac{3}{7} \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1&0&0&\frac{5}{7}&-\frac{2}{7}&-\frac{1}{7} \\ 0&1&0&-\frac{2}{7}&\frac{5}{7}&-\frac{1}{7} \\ 0&0&1&-\frac{1}{7}&-\frac{1}{7}&\frac{3}{7} \end{array}\right]$

$\underbrace{\begin{bmatrix} \frac{5}{7}&-\frac{2}{7}&-\frac{1}{7} \\ -\frac{2}{7}&\frac{5}{7}&-\frac{1}{7} \\ -\frac{1}{7}&-\frac{1}{7}&\frac{3}{7} \end{bmatrix}}_{A^{-1}}$

(3) $\det(A) = 0 \Leftrightarrow A^{-1}$ 존재 X 즉, (b) 성립 X

1-2

(1) $a \to b$ $\quad Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b$ 역행렬은 유일하므로 유일해 존재

$b \to c$ $\quad b=0$ 일때 $x = A^{-1}\cdot 0 = 0$ 을 가짐

$c \to a$ $\quad Ax=0$ 의 해가 자명해 $\Leftrightarrow$ A의 열벡터 선형독립 $\Leftrightarrow \det(A) \neq 0 \Leftrightarrow$ A는 가역

(2) $Ax = b \Leftrightarrow A^{-1}Ax = A^{-1}b \Leftrightarrow x = A^{-1}b$

## 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 $B^TB, BB^T$ 를 계산하고, 이들의 고유값을 계산해 특이값 $\sigma$와 고유벡터를 통해 직교대각화하여 $B^TB = VDV^T, BB^T = UDU^T$ 를 구합니다.
2) $0$이 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.

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

$$BB' = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \qquad \lambda = 2, 1 \qquad \Sigma = \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \end{pmatrix}$$

$$\lambda = 2 \qquad \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\overset{x}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} = 2\overset{x}{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}} \Rightarrow \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

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

$$\lambda = 1 \qquad \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$B'B = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix} \Rightarrow (1-\lambda)\cdot(1-\lambda)^2 - (1-\lambda) = (1-\lambda)(\lambda^2 - 2\lambda)$$

$$= \lambda(\lambda-2)(1-\lambda) = 0$$

$$\lambda = 0, 1, 2$$

$$\lambda = 2 \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 2\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \begin{matrix} x_1 + x_2 = 2x_1 \\ x_1 + x_2 = 2x_2 \\ x_3 = 2x_3 \end{matrix} \Rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\lambda = 1 \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \qquad \begin{matrix} x_1 + x_2 = x_1 \\ x_1 + x_2 = x_2 \\ x_3 = x_3 \end{matrix} \Rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\lambda = 0 \qquad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \qquad V' = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

# 문제 3　Convex Sets & Functions

## 3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

$$C_1 = \left\{(x, y) \in \mathbb{R}^2 : x + y \leq 1\right\}$$
$$C_2 = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$$
$$C_3 = \left\{(x, y) \in \mathbb{R}^2 : y \geq e^x\right\}$$

2. 함수 $f : \mathbb{R}^n \to \mathbb{R}$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

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

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

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

2. Convex Optimization

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

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

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

3-1

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

$\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)$

$= (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) = (x_3, y_3)$

$x_3 + y_3 = \lambda(x_1 + y_1) + (1-\lambda)(x_2 + y_2) \le \lambda + (1-\lambda) = 1$

$\therefore \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) \in C_1 \Rightarrow C_1$ is Convex

$x_1, x_2 \in C_2 \Rightarrow \|x_1\|_1 \le 1 \quad , \quad \|x_2\|_1 \le 1$

$\|\lambda x_1 + (1-\lambda)x_2\| \le \lambda\|x_1\|_1 + (1-\lambda)\|x_2\|_1 \le \lambda + (1-\lambda) = 1$

$\therefore \lambda x_1 + (1-\lambda)x_2 \in C_2 \Rightarrow C_2$ is Convex

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

$\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)$

$= (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) = (x_3, y_3)$

$y_3 = \lambda y_1 + (1-\lambda)y_2 \ge \lambda e^{x_1} + (1-\lambda)e^{x_2} \ge e^{\lambda x_1 + (1-\lambda)x_2} = e^{x_3}$

$\therefore \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) \in C_3 \Rightarrow C_3$ is Convex

3-1

(2) $(x_1, t_1), (x_2, t_2) \in S \Rightarrow t_1 \ge f(x_1), \quad t_2 \ge f(x_2)$

$f$는 Convex function 이므로

$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2) \le \lambda t_1 + (1-\lambda)t_2$

$\therefore (\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2) \in S$

3-2

(1)

(a) let, $f(x)+g(x)=h(x)$

$$
\begin{array}{rl}
 & f(\lambda x_1+(1-\lambda)x_2) \le \lambda f(x_1)+(1-\lambda)f(x_2) \\
+) & g(\lambda x_1+(1-\lambda)x_2) \le \lambda g(x_1)+(1-\lambda)g(x_2) \\
\hline
 & h(\lambda x_1+(1-\lambda)x_2) \le \lambda h(x_1)+(1-\lambda)h(x_2)
\end{array}
$$

$\therefore$ $h(x)=f(x)+g(x)$ is convex function

(b) $A(\lambda x_1+(1-\lambda)x_2)+b = \lambda(Ax_1+b)+(1-\lambda)(Ax_2+b)$

$f(A(\lambda x_1+(1-\lambda)x_2)+b) \le \lambda f(Ax_1+b)+(1-\lambda)f(Ax_2+b)$

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

(2) $A(\lambda x_1+(1-\lambda)x_2)-b = \lambda(Ax_1-b)+(1-\lambda)(Ax_2-b)$

$\|A(\lambda x_1+(1-\lambda)x_2)-b\|^2 = \|\lambda(Ax_1-b)+(1-\lambda)(Ax_2-b)\|^2 \le \lambda\|Ax_1-b\|^2+(1-\lambda)\|Ax_2-b\|^2$

즉, $f(\lambda x_1+(1-\lambda)x_2) \le \lambda f(x_1)+(1-\lambda)f(x_2)$ $\therefore$ $f(x)$ is convex function

# 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.

## 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**

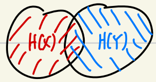
- Elements of Information Theory (2nd Ed.), T. M. Cover & J. A. Thomas
- Convex Optimization for ML, Prof. Changho Suh

Data Science Lab
담당자: 14 기 박창용, 어희정, 여준호
qkrckddyd0@yonsei.ac.kr
heejung.uh@gmail.com
asap03153@gmail.com

# 4-1

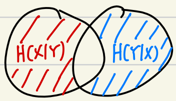
(a) $H(X) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3}$

$H(Y) = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}$

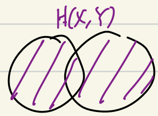



(b) $H(X|Y) = \frac{1}{3}H(X|Y=0) + \frac{2}{3}H(X|Y=1) = \frac{1}{3}(-1\log 1 - 0\log 0) + \frac{2}{3}(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}) = \frac{1}{3}\cdot 0 + \frac{2}{3}\cdot 1 = \frac{2}{3}$

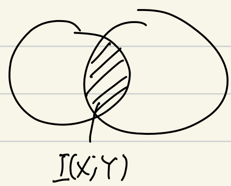
$H(Y|X) = \frac{2}{3}H(Y|X=0) + \frac{1}{3}H(Y|X=1) = \frac{2}{3}(-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2}) + \frac{1}{3}(-0\log 0 - 1\log 1) = \frac{2}{3}\cdot 1 + \frac{1}{3}\cdot 0 = \frac{2}{3}$




(c) $H(X,Y) = -\frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} - \frac{1}{3}\log\frac{1}{3} = \log 3$




(d) (e) $I(X;Y) = H(Y) - H(Y|X)$




# 4-2

(a) $p = (0.4, 0.6)$ $q = (0.5, 0.5)$ $D(p\|q) = 0.4\log\frac{0.4}{0.5} + 0.6\log\frac{0.6}{0.5}$

$D(q\|p) = 0.5\log\frac{0.5}{0.4} + 0.5\log\frac{0.5}{0.6}$

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

(b) $D(p\|q) = \sum p(x)\log\frac{p(x)}{q(x)} = \sum -p(x)\log\frac{q(x)}{p(x)}$

By Jenson's Inequality $D(p\|q) = \sum -p(x)\log\frac{q(x)}{p(x)} \geq -\log\left(\sum p(x)\frac{q(x)}{p(x)}\right) = -\log 1 = 0$