기수 : 15기 　 이름 : 이지원

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 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.

$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$

$\Delta_1 = 2 > 0$

$\Delta_2 = \det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 3 > 0$

$\Delta_3 = 7 > 0$

∴ A는 positive definite.

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

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

$A^{-1} = \begin{pmatrix} 5/7 & -2/7 & -1/7 \\ -2/7 & 5/7 & -1/7 \\ -1/7 & -1/7 & 3/7 \end{pmatrix} = \frac{1}{7}\begin{pmatrix} 5 & -2 & -1 \\ -2 & 5 & -1 \\ -1 & -1 & 3 \end{pmatrix}$

(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.

$xy \to kxy$로 변경 $\begin{pmatrix} 2 & k & 1 \\ k & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix} \to \det(A) = 10 - k(3k-1) + k - 2$

$= 8 - 3k^2 + 2k = 0$

$3k^2 - 2k - 8 = 0$ 　 $(k-2)(3k+4) = 0$

$k = 2$ 또는 $k = -\frac{4}{3}$

∴ $\det(A) = 0$이면 A는 non-invertible 하므로 $Ax = b$는 유일해를 갖지 않는다.

1-2. (Optional)

(1) (a), (b), (c) 세 조건들이 동치임을 증명하시오.

i) (a) → (b) : $Ax = b$에 대해

$A^{-1}Ax = A^{-1}b \Rightarrow Ix = A^{-1}b \Rightarrow x = A^{-1}b$

∴ 해가 존재하고 유일하다.

ii) (b) → (c)

$Ax = b$. 모든 b에 대해서 유일.

$Ax = 0$, 　 $x = 0$.

trivial solution을 갖는다.

iii) (c) → (a)

$Ax = 0 \to x = 0$

$\ker(A) = 0$. A가 정사각행렬이므로

A는 full rank $\Leftrightarrow \det(A) \neq 0 \Leftrightarrow$ invertible. ∎

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

A가 invertible $\Rightarrow A^{-1}$가 존재.

임의의 b에 대해 $x = A^{-1}b$.

$Ax = A(A^{-1}b) = (AA^{-1})b = b$

모든 b에 대해 $x = A^{-1}b$가 존재한다.

∴ A가 가역이면 임의의 b에 대해 $Ax = b$의 해 x가 항상 존재한다.

→ A가 가역이라는 것은 $T(x) = Ax$가 역함수를 가지는 변환. 어떤 벡터 b든, x가 반드시 존재한다 (onto)

기수 : 15기 이름 : 이지원

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.

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

① $BB^T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$

$\begin{vmatrix} 2-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} = (2-\lambda)(1-\lambda) = 0$

$\lambda = 2, 1$

i) $\lambda = 2$ $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0$ $\therefore \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

ii) $\lambda_2 = 1$ $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0$ $\therefore \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

$\Rightarrow U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ $\sigma_1 = \sqrt{2}$ $\sigma_2 = 1$

② $B^TB = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

$\begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)\{(1-\lambda)^2 - 1\}$

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

$\lambda = 2, 1, 0$

i) $\lambda = 2$. $\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0$

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

ii) $\lambda = 1$ $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0$ $\therefore \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

iii) $\lambda = 0$ $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = 0$ $\therefore \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

$\Rightarrow V = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$ $V^T = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix}$

$\Sigma = \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$

$B = U\Sigma V^T$

$= \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{U}\underbrace{\begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}}_{\Sigma}\underbrace{\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{pmatrix}}_{V^T}$

기수 : 15기 이름 : 이지원

문제 3 Convex Sets & Functions

3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

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

· $C_1 = \{(x,y) \in \mathbb{R}^2 \mid x+y \leq 1\}$
임의의 $(x_1,y_1)(x_2,y_2) \in C_1$. $\lambda \in [0,1]$
$x_1+y_1 \leq 1$, $x_2+y_2 \leq 1$
$(x,y) = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2)$
$x+y = \lambda(x_1+y_1) + (1-\lambda)(x_2+y_2) \leq \lambda \cdot 1 + (1-\lambda) \cdot 1 = 1$
$\therefore$ $(x,y) \in C_1$. 이므로 C_1은 Convex set.

· $C_2 = \{x \in \mathbb{R}^n \mid \|x\|_1 \leq 1\}$
$x_1, x_2 \in C_2$, $\lambda \in [0,1]$
$\|\lambda x_1 + (1-\lambda)x_2\|_1 \leq \|\lambda x_1\|_1 + \|(1-\lambda)x_2\|_1$
$= \lambda\|x_1\|_1 + (1-\lambda)\|x_2\|_1 = \lambda + (1-\lambda) = 1$
$\therefore$ $\lambda x_1 + (1-\lambda)x_2 \in C_2$. $\therefore$ C_2는 Convex set.

· $C_3 = \{(x,y) \in \mathbb{R}^2 \mid y \geq e^x\}$
$(x_1,y_1), (x_2,y_2) \in C_3$, $y_1 \geq e^{x_1}$, $y_2 \geq e^{x_2}$
$x = \lambda x_1 + (1-\lambda)x_2$, $y = \lambda y_1 + (1-\lambda)y_2$
$y = \lambda y_1 + (1-\lambda)y_2 \geq \lambda e^{x_1} + (1-\lambda)e^{x_2}$
e^x는 convex function이므로.
$e^x = e^{\lambda x_1 + (1-\lambda)x_2} \leq \lambda e^{x_1} + (1-\lambda)e^{x_2}$
$y \geq \lambda e^{x_1} + (1-\lambda)e^{x_2} \geq e^x$
$\therefore$ $(x,y) \in C_3$. $\therefore$ C_3는 Convex set.

2. 함수 $f : \mathbb{R}^n \to \mathbb{R}$가 convex function일 때, f의 epigraph인 집합 S가 convex set임을 보이시오.

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

$(x_1, t_1), (x_2, t_2) \in S$를 잡으면,
$t_1 \geq f(x_1)$, $t_2 \geq f(x_2)$
$\lambda \in [0,1]$에 대해. $(x,y) = (\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2)$
f의 Convex function 이므로
$f(x) = f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \leq \lambda t_1 + (1-\lambda)t_2 = t$
$\therefore$ $t \geq f(x)$ 이므로 $(x,y) \in S$. S는 Convex set ■

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.

$(f+g)(\lambda x + (1-\lambda)y) = f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) + \lambda g(x) + (1-\lambda)g(y) = \lambda(f+g)(x) + (1-\lambda)(f+g)(y)$
$\therefore$ $f(x)+g(x)$ 또는 Convex.

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

$x_1, x_2, \lambda \in [0,1]$.
$h(\lambda x_1 + (1-\lambda)x_2) = f(A(\lambda x_1 + (1-\lambda)x_2) + b)$
$= f(\lambda(Ax_1+b) + (1-\lambda)(Ax_2+b)) \leq \lambda f(Ax_1+b) + (1-\lambda)f(Ax_2+b) = \lambda h(x_1) + (1-\lambda)h(x_2)$ $\therefore$ $h(x) = f(Ax+b)$는 Convex.

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.

↳ $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$.

$u = Ax_1 - b$. $v = Ax_2 - b$

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

$\|\lambda u + (1-\lambda)v\|^2 = \lambda^2\|u\|^2 + (1-\lambda)^2\|v\|^2 + 2\lambda(1-\lambda)u^T v$
$\|u-v\|^2 \geq 0$. $\|u-v\|^2 = \|u\|^2 + \|v\|^2 - 2u^T v \geq 0$. $\Rightarrow$ $2u^T v \leq \|u\|^2 + \|v\|^2$
$\|\lambda u + (1-\lambda)v\|^2 \leq \lambda^2\|u\|^2 + (1-\lambda)^2\|v\|^2 + \lambda(1-\lambda)(\|u\|^2 + \|v\|^2)$,
$= (\lambda^2 + \lambda(1-\lambda))\|u\|^2 + ((1-\lambda)^2 + \lambda(1-\lambda))\|v\|^2$
$= \lambda\|u\|^2 + (1-\lambda)\|v\|^2$

$\therefore$ $f(\lambda x_1 + (1-\lambda)x_2) = \|\lambda u + (1-\lambda)v\|^2 \leq \lambda\|u\|^2 + (1-\lambda)\|v\|^2$
$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$
$f(x) = \|Ax - b\|^2$는 Convex.

기수 : 15기 이름 : 이지원

문제 4 정보이론 (Information Theory)

4-1. Entropy

확률분포 $p(x, y)$가 다음과 같이 주어졌다고 합시다.

$X \backslash Y$	0	1	
0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{2}{3}$
1	0	$\frac{1}{3}$	$\frac{1}{3}$
	$\frac{1}{3}$	$\frac{2}{3}$	

다음 값들을 구하시오.

(a) $H(X), H(Y)$. $H(X) = -\sum_{x} P(x)\log_2 P(x) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}$

$H(Y) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

(b) $H(X \mid Y), H(Y \mid X)$. $Y=0 \rightarrow P(X=0|Y=0)=\frac{1}{3}$. $P(X=1|Y=0)=0 \rightarrow H(X|Y=0)=0$

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

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

$\therefore \frac{2}{3}$

$X=0 \rightarrow P(Y=0|X=0)=\frac{1}{2}$. $P(Y=1|X=0)=\frac{1}{2} \rightarrow H(Y|X=0)=1$

$X=1 \rightarrow P(Y=1|X=1)=1 \Rightarrow H(Y|X=1)=0$.

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

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

$-\sum_{(x,y)} P(x,y)\log_2 P(x,y) = -3\cdot\frac{1}{3}\log_2\frac{1}{3} = \log_2 3$

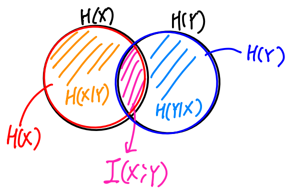
(d) $H(Y) - H(Y \mid X)$.

$\log_2 3 - \frac{2}{3} - \frac{2}{3} = \log_2 3 - \frac{4}{3}$

(e) $I(X; Y)$. $= H(Y) - H(Y|X) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)$

$= \log_2 3 - \frac{4}{3}$

(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.

이산분포 $p=(0.9, 0.1)$ $q=(0.5, 0.5)$ $D(p\|q) = 0.9\ln\frac{0.9}{0.5} + 0.1\ln\frac{0.1}{0.5}$

$D(q\|p) = 0.5\ln\frac{0.5}{0.9} + 0.5\ln\frac{0.5}{0.1}$ $\neq$.

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

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

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

$r(x) = \frac{q(x)}{p(x)}$. $(p(x)>0)$ 라 하자. $D(p\|q) = -E_p[\log r(x)]$

by Jensen 부등식. $E_p[\log r(x)] \leq \log(E_p(r(x)))$

$E_p(r(x)) = \sum_{x} p(x)\frac{q(x)}{p(x)} = \sum_{x} q(x) = 1$ 이므로

$E_p[\log r(x)] \leq \log 1 = 0 \rightarrow -E_p[\log r(x)] \geq 0$

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

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