기수 : 14기 이름 : 김서윤

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.

$f_x = 2x+y+z$
$f_y = 2y+x+z$
$f_z = 3z+y+x$

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

$D_1 = \det(2) = 2$

$D_2 = \det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 4-1 = 3$

$D_3 = \det\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix} = 2\times5-2-1 = 7$

∴ PD 임.

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

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

(3) 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.

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

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

det (A) = 0 이라면, A는 비가역적.
∴ 조건(a) 성립x. 동치이기 때문에 조건 (b)도 성립x.

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.

기수 : 14기 이름 : 김서원

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 σ 와 고유벡터를 통해 직교대각화하여 $B^TB = VDV^T, BB^T = UDU^T$ 를 구합니다.
> 2) 0 이 아닌 특이값들을 내림차순으로 나열하여 Σ 를 구성하고, 이들을 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.

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

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

i) U 값 : BB^T 의 단위고유벡터

$\det(BB^T - \lambda I) = 0$

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

$\det(BB^T - \lambda I) = (2-\lambda)(1-\lambda) = 0$

$\therefore \lambda = 2, 1$

ii) Σ : $\sqrt{2}, 1$

① $\lambda = 2$ 인 경우,

$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = 2 \begin{pmatrix} a \\ b \end{pmatrix}$

$\begin{pmatrix} 2a \\ b \end{pmatrix} = \begin{pmatrix} 2a \\ 2b \end{pmatrix}$

$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

② $\lambda = 1$ 인 경우,

$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$

$\begin{pmatrix} 2a \\ b \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}$ $v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

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

iii) V 값 : B^TB 의 단위 고유 벡터

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

$\det(B^TB - \lambda I) = 0$

$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{pmatrix}$

$(1-\lambda)\left((1-\lambda)^2 - 1\right)$

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

i) $\lambda = 0$

$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$ $a+b=0$, $a+b=0$, $c=0$

$v = \begin{pmatrix} a \\ -a \\ 0 \end{pmatrix}$ $a=1$ 설정

$v = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$

ii) $\lambda = 1$

$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$ $a=0$, $b=0$

$v = \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix}$ $c=1$ 설정

$v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

iii) $\lambda = 2$

$\begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = 0$ $-a+b=0$, $a-b=0$, $-c=0$, $a=b$ $c=0$.

$v = \begin{pmatrix} a \\ a \\ 0 \end{pmatrix}$ $a=1$, $v = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

기수 : 14기   이름 : 김서윤

문제 3 Convex Sets & Functions

3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

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

① C_1 (x_1, y_1) $(x_2, y_2) \in C_1$

$x_1 + y_1 \le 1$, $x_2 + y_2 \le 1$

$\lambda \in [0,1]$

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

$x_\lambda + y_\lambda = \lambda(x_1+y_1) + (1-\lambda)(x_2+y_2) \le \lambda + 1 - \lambda \le 1$

∴ Convex set.

② C_2

$x_1, x_2 \in C_2$ $\|x_1\| \le 1$ $\|x_2\| \le 1$

$\lambda \in [0,1]$ $x_\lambda = \lambda x_1 + (1-\lambda)x_2$

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

∴ Convex set

③ C_3

$y_1 \ge e^{x_1}$ $y_2 \ge e^{x_2}$

$\lambda \in [0,1]$ ∴ $(x_\lambda, y_\lambda) = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2)$

e^x는 convex.

∴ $e^{x_\lambda} = e^{\lambda x_1 + (1-\lambda)x_2} \le \lambda e^{x_1} + (1-\lambda)e^{x_2}$

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

∴ $y_\lambda \ge \lambda e^{x_1} + (1-\lambda)e^{x_2} \ge e^{x_1}$

∴ Convex set

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

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

(x_1, t_1) $(x_2, t_2) \in S$

$t_1 \ge f(x_1)$, $t_2 \ge f(x_2)$

$\lambda \in [0,1]$

$(x_\lambda, t_\lambda) = (\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2)$

$t_\lambda \ge x_\lambda$

f가 convex이기 때문에

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

∴ $t_\lambda \ge \lambda f(x_1) + (1-\lambda)f(x_2) \ge f(x_\lambda)$

∴ $(x_\lambda, t_\lambda) \in S$

∴ Convex.

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.

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

∴ Convex.

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

$f(Ax+b) = h(x)$라 하자.

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

∴ $h(\lambda x_1 + (1-\lambda)x_2) = f($ 〃 $)$

f(·)가 convex이므로,

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

$t_1 = Ax_1 + b$ $t_2 = Ax_2 + b$

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

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

∴ $f(Ax+b)$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.

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

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

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

차이 $\lambda\|u\|^2 + (1-\lambda)\|v\|^2 - \|\lambda u + (1-\lambda)v\|^2$

$= \lambda\|u\|^2 + (1-\lambda)\|v\|^2 - (\lambda^2\|u\|^2 + (1-\lambda)^2\|v\|^2 + 2\lambda(1-\lambda)u^T v)$

$= \lambda(1-\lambda)\|u\|^2 + \lambda(1-\lambda)\|v\|^2 - 2\lambda(1-\lambda)u^T v$

$= \lambda(1-\lambda)\|u - v\|^2 \ge 0$

$\|\lambda u + (1-\lambda)v\|^2 \le \lambda\|u\|^2 + (1-\lambda)\|v\|^2$

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

∴ $f(x)$는 convex.

기수 : 14기 이름 : 김지윤

문제 4 정보이론 (Information Theory)

$H(X|Y)$
$Y=0$ 일때.
$P(X=0, Y=0)=1 \Rightarrow H(X|Y=0)=0$
$Y=1$ 일때
$P(X=0, Y=1) = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}$
$P(X=1, Y=1) = \frac{1}{2}$

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

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

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

(b) $H(X \mid Y), H(Y \mid X)$. $* \; H(X|Y) = \sum_y P(Y=y)\, H(X|Y)$

$= P(Y=1) H(X|Y=1) = \frac{2}{3} \times (-\log\frac{1}{2})$

$* \; H(Y|X) = P(X=0) H(Y|X=0)$

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

(c) $H(X, Y)$. $= H(Y) + H(X|Y)$

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

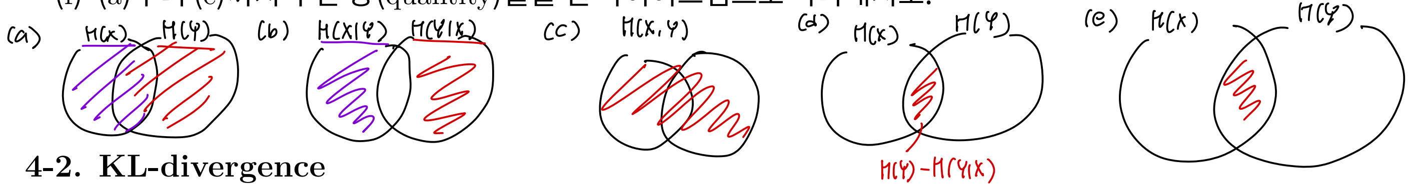
(d) $H(Y) - H(Y \mid X)$.

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

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

∴ (d)와 동일.

(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.

$D(q\|p) \neq D(p\|q)$ not symmetric → KL Divergence 특징

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

$D(p\|q) = E_p \log\frac{p(x)}{q(x)} = E_p \log\left(-\underbrace{\frac{q(x)}{p(x)}}_{z}\right)$

$D(p\|q) = E_p(-\log z)$

Jensen's Inequality에 의해서.

$E(-\log z) \geq -\log(E(z))$

$E(z) = \sum_x p(x)\frac{q(x)}{p(x)} = \sum_x q(x) = 1$

$\therefore E(-\log z) \geq -\log 1$

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

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