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.

$\frac{\partial^2 f}{\partial x^2} = 2, \ \frac{\partial^2 f}{\partial y^2} = 2, \ \frac{\partial^2 f}{\partial z^2} = 3$

1-1. 다음과 같은 삼변수 함수 $f(x, y, z)$가 주어져 있습니다.

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

$\frac{\partial^2 f}{\partial x \partial y} = 1, \ \frac{\partial^2 f}{\partial y \partial z} = 1, \ \frac{\partial^2 f}{\partial z \partial x} = 1$

(1) 이 함수의 헤시안 행렬(Hessian Matrix) A를 구하고 PD(Positive Definite), PSD, ND, NSD 중 무엇인지 판별하시오.

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

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

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

(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.

$A(k) = \begin{pmatrix} 2 & k & 1 \\ k & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$, $\det(A(k)) = 0$ 이면 $A(k)$는 특이행렬 → 역행렬 X → 조건(b) 성립 X

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 : 신지원

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

$\sigma_1 = \sqrt{2}, \quad \sigma_2 = 1$

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

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

$\lambda = 2 \ ; \quad v_1 = \frac{1}{\sqrt{2}}(1, 1, 0)^T$

$\lambda = 1 \ ; \quad v_2 = (0, 0, 1)^T$

$\lambda = 0 \ ; \quad v_3 = \frac{1}{\sqrt{2}}(-1, 1, 0)^T$

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

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

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

기수 : 14기　이름 : 신지원

문제 3 Convex Sets & Functions

3-1. Convex Set

집합 C내의 임의의 두점 x_1, x_2 와 $0 \le \theta \le 1$ 일 때,
$\theta x_1 + (1-\theta)x_2 = C$ 이면 C는 convex set

1. 다음 집합들이 Convex set인지 증명하시오.

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

C_1)
$u = (x_1, y_1),\ v = (x_2, y_2) \in C_1$
$x_1 + y_1 \le 1,$
$x_2 + y_2 \le 1$
$\theta(x_1 + y_1) + (1-\theta)(x_2 + y_2) \le \theta \cdot 1 + (1-\theta)\cdot 1 = 1$
$\therefore \theta u + (1-\theta)v \in C_1 \rightarrow$ convex

C_2) $x, y \in C_2$ 가정
$\|x\|_1 \le 1,\ \|y\|_1 \le 1,$ $\|\theta x + (1-\theta)y\|_1 \le \|\theta x\|_1 + \|(1-\theta)y\|_1$
$= \theta\|x\|_1 + (1-\theta)1,$
$\theta(1) + (1-\theta)(1) = 1$
$\therefore$ convex

C_3) $f(x) = e^x$
$f''(x) = e^x > 0$ 이므로
epi그래프인 C_3은 convex

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

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

$(x_1, t_1), (x_2, t_2) \in S$
$t_1 \ge f(x_1),\ t_2 \ge f(x_2)$
$\theta t_1 + (1-\theta)t_2 \ge \theta f(x_1) + (1-\theta)f(x_2) \ge f(\theta x_1 + (1-\theta)x_2)$
$(\theta x_1 + (1-\theta)x_2,\ \theta t_1 + (1-\theta)t_2) \in S$
$\therefore$ 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, $\underbrace{f(x) + g(x)}_{h(x)}$ 또한 convex임을 보이시오.

$h(\theta x + (1-\theta)y) = f(\theta x + (1-\theta)y) + g(\theta x + (1-\theta)y) \le [\theta f(x) + (1-\theta)f(y)] + [\theta g(x) + (1-\theta)g(y)]$
$= \theta[f(x) + g(x)] + (1-\theta)[f(y) + g(y)]$
$= \theta h(x) + (1-\theta)h(y)$
$\therefore$ convex

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

$h(x) = f(Ax + b)$
$h(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + \underbrace{b}_{\theta b + (1-\theta)b})$
$f(\theta(Ax + b) + (1-\theta)(Ay + b))$
$\le \theta f(Ax + b) + (1-\theta)f(Ay + b) = \theta h(x) + (1-\theta)h(y)$
$\therefore$ 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.

$f(\theta x + (1-\theta)y) = \|A(\theta x + (1-\theta)y) - b\|^2$
$= \|\theta(Ax - b) + (1-\theta)(Ay - b)\|^2$

$u = Ax - b$
$v = Ay - b$
$\|\theta u + (1-\theta)v\|^2 \le \theta\|u\|^2 + (1-\theta)\|v\|^2$
$= \theta\|Ax - b\|^2 + (1-\theta)\|Ay - b\|^2$
$= \theta f(x) + (1-\theta)f(y)$
$\therefore$ f는 convex

기수 : 14기   이름 : 선지원

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	$\frac{1}{3}$

$p(0,0)=\frac{1}{3}$ $p(x=0)=\frac{2}{3}$
$p(0,1)=\frac{1}{3}$ $p(x=1)=\frac{1}{3}$
$p(1,0)=0$ $p(y=0)=\frac{1}{3}$
$p(1,1)=\frac{1}{3}$ $p(y=1)=\frac{2}{3}$

다음 값들을 구하시오.

(a) $H(X), H(Y)$. $H(X)=\log_2 3-\frac{2}{3}=0.918$
$H(Y)=\log_2 3-\frac{2}{3}=0.918$

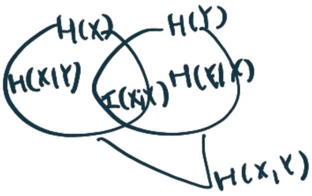
(b) $H(X \mid Y), H(Y \mid X)$.
$H(X,Y)-H(Y)=1.585-0.918=0.667$
$H(X,Y)-H(X)=1.585-0.918=0.667$

(c) $H(X,Y)$.
$=\log_2 3$

(d) $H(Y)-H(Y \mid X)$. $\log_2 3-\frac{2}{3}-\frac{2}{3}=0.251$

(e) $I(X;Y)$. 0.251

(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.
ex) $p=\{0.5, 0.5\}$
$q=\{0.1, 0.9\}$
$D(q\|p)=\sum q_i \log\frac{q_i}{p_i}=0.1\cdot\log\frac{0.1}{0.5}+0.9\log\frac{0.9}{0.5} \approx 0.533$

(b) $D(p\|q)\geq 0$임을 보이시오. (Hint : Jensen's Inequality)
$D(p\|q)=0.5\cdot\log\frac{0.5}{0.1}+0.5\log\frac{0.5}{0.9}$
≈ 0.735
$\therefore D(p\|q)\neq D(q\|p)$

$-D(p\|q)=\sum p(x)\log\frac{q(x)}{p(x)} \leq \log\left(\sum p(x)\frac{q(x)}{p(x)}\right)$
$\underbrace{\quad}_{=1}$ $=0$

$\therefore -D(p\|q)\leq 0,$
$D(p\|q)\geq 0$ "

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