기수 : 14기 이름 : 윤동영

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.

$f_x = 2x+y+z$, $f_y = x+2y+z$, $f_z = x+y+3z$. $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$

$D_1 = 2 > 0$.
$D_2 = \det\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 4-1 = 3 > 0$ A는 PD.
$D_3 = \det(A) = 10-2-1 = 7 > 0$.

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

$\det(A) = 7$. $C_{11} = +(6-1) = 5$ $C_{12} = -(3-1) = -2$ $C_{13} = +(1-2) = -1$
$C_{21} = -(3-1) = -2$ $C_{22} = +(6-1) = 5$ $C_{23} = -(2-1) = -1$
$C_{31} = +(1-2) = -1$ $C_{32} = -(2-1) = -1$. $C_{33} = +(4-1) = 3$

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

(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.

성립X. 연립방정식이 유일한 해를 갖는다는 것은 A가 inversible 해야 하는데, $\det(A) \neq 0$ 과 동치이다.
$\therefore \det(A) = 0$ 이면 역행렬이 존재하지 않으므로 유일한 해가 존재X.

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.

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 σ와 고유 벡터를 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^T 를 각각 구하시오.

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

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

$\lambda_1 = 2.\ \lambda_2 = 1. \to \sigma_1 = \sqrt{2}.\ \sigma_2 = 1.\quad u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\ u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.\quad U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

ii) $B^TB = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\lambda_1 = 2.\ \lambda_2 = 1.\ \lambda_3 = 0.$

$\lambda_1 = 2$인 경우. $\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} x = 0 \quad \therefore x_1 = x_2, x_3 = 0.\quad v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$\lambda_2 = 1$인 경우. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} x = 0. \quad \therefore x_1 = 0.\ x_2 = 0.\ x_3 = 1.\quad v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$\lambda_3 = 0$인 경우. $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} x = 0. \quad \therefore x_1 = -x_2, x_3 = 0\quad v_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

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

기수 : 14기 이름 : 윤동영

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 : $x+y \le 1$를 만족하면 두 점을 이은 선분도 포함되므로 convex.

C_2 : $f(x) = \|x\|_1$ (norm func.)은 convex이므로 subset $\|x\|_1 \le 1$ 도 convex.

C_3 : $f(x) = e^x$. $f''(x) = e^x > 0$ 이므로 convex.

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

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

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

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

θ로 내분하는 내분점. $\theta x_1 + (1-\theta) x_2$.

f가 convex이므로 $f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2)$

$\rightarrow$ $f(\theta x_1 + (1-\theta)x_2) \le \theta t_1 + (1-\theta) t_2$.

$\therefore$ 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.

$h(x) = f(x) + g(x)$ $h''(x) = f''(x) + g''(x)$ $h''(x) \ge 0$ $\because$ f, g가 convex. h도 convex.

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

affine 변환 $\rightarrow$ $h(x) = f(Ax+b)$. x, y 내분점으로 보면 $\rightarrow$ $A(\theta x + (1-\theta) y) + b = \theta(Ax+b) + (1-\theta)(Ay+b)$.

f가 convex이므로, 위 식도 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.

$g(x) = \|x\|^2$, $h(x) = Ax - b$

affine 변환은 convex를 그대로 보존시키고, L2 norm function의 제곱은 convex하므로

$f(x) = g(h(x))$도 convex

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

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

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

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

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

$H(X,Y) = \sum_{x,y} -p(x,y)\log p(x,y) = -3\cdot(\frac{1}{3}\log_2\frac{1}{3}) = \log_2 3$

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

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

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

$I(X;Y) = H(X) + H(Y) - H(X,Y) = (-\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}) \times 2 - \log_2 3$

(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.

P와 q의 분포가 다른 경우.

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

$-\sum p(x)\log\frac{q(x)}{p(x)}$ 에서 $-\log$는 convex이므로,

Jensen 부등식 적용하면

$\sum p(x)\log\frac{q(x)}{p(x)} \leq \log(\sum p(x)\frac{q(x)}{p(x)})$

↳ $\log 1 = 0$.

$\therefore D(p\|q) = -\sum p(x)\log\frac{q(x)}{p(x)} \geq 0$

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