기수 : 15기 이름 : 안재민

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.

$H(f) = \begin{bmatrix} \partial x^2 & \partial x \partial y & \partial x \partial z \\ \partial y \partial x & \partial y^2 & \partial y \partial z \\ \partial z \partial x & \partial z \partial y & \partial z^2 \end{bmatrix}$ $f_x = 2x+y+z$ $f_z = x+y+3z$ $f_y = x+2y+z$ $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$, $\lambda_i > 0$ 이므로 PD

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

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

(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.

행렬 A가 가역적 $\longleftrightarrow$ $\det(A) \neq 0$ 이므로 해가 무수히 많거나 존재하지 않는다

1-2. (Optional)

(1) (a), (b), (c) 세 조건들이 동치임을 증명하시오.

$a \to b$ $Ax = b$ $x = A^{-1}b$. $b \to c$ $b = 0$ $Ax = 0$ 유일 해 $c \to a$ $Ax = 0$ 유일해 $\to$ linear independent $\det(A) \neq 0$

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

A^{-1} 존재 $Ax = b \to A^{-1}(Ax) = A^{-1}b$ $\therefore x = A^{-1}b$

기수 : 15기 이름 : 안재민

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 U : BB^T의 단위고유벡터

② Σ : 특이값을 주대각원소로 가지는 대각행렬

③ V : A^TA의 단위고유벡터

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

① $BB^T = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ $\det(BB^T - \lambda I) = 0$ $\det\left(\begin{bmatrix} 2-\lambda & 0 \\ 0 & 1-\lambda \end{bmatrix}\right) = 0$

$\lambda = 1, 2$ $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

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

③ $B^TB = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\det\left(\begin{bmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{bmatrix}\right) = 0$ $\lambda = 2, 1, 0$

$V = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$

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

기수 : 15기 이름 : 안재민

문제 3 Convex Sets & Functions

3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

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

$C_1 : (x_1, y_1), (x_2, y_2) \in C_1$

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

$C_2 : \|\lambda x_1 + (1-\lambda)x_2\|_1 \le \lambda\|x_1\|_1 + (1-\lambda)\|x_2\|_1 \le 1$

$C_3 : \lambda y_1 + (1-\lambda)y_2 \ge \lambda e^{x_1} + (1-\lambda)e^{x_2} \ge e^{\lambda x_1} + e^{(1-\lambda)x_2}$

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

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

$P_1(x_1, t_1), P_2(x_2, t_2) \subset S$ f는 convex function이므로 $f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2) \le \lambda t_1 + (1-\lambda)t_2$

$P_\lambda = \lambda P_1 + (1-\lambda)P_2$ $\therefore P_\lambda$도 $t \ge f(x)$ 만족하므로 S는 convex set임.

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.

$x_1, x_2 \in R^d$, $0 \le \lambda \le 1$

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

$(\because f(x), g(x)$ is convex function$)$

$f(x) + g(x)$ is convex

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

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

$(\because f(x)$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.

$x_1, x_2 \in R^d$, $0 \le \lambda \le 1$

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

$= \|\lambda u + (1-\lambda)v\|^2 \quad (u = Ax_1 - b \quad v = Ax_2 - b)$

$g(x) = \|x\|^2$라 할때 $g(x)$는 convex이므로

$\|\lambda u + (1-\lambda)v\|^2 \le \lambda\|u\|^2 + (1-\lambda)\|v\|^2 = \lambda\|Ax_1 - b\|^2 + (1-\lambda)\|Ax_2 - b\|^2$

$\therefore f(x)$ is convex function.

기수 : 15기 이름 : 안재민

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

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

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

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

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

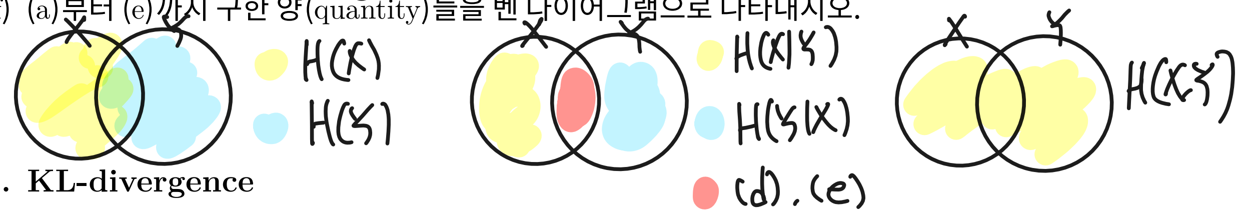
$H(X,Y) = H(X) + H(Y|X) = \log_2 3$

(d) $H(Y) - H(Y \mid X)$. $\log_2 3 - \frac{4}{3}$

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

$H(X) + H(Y) - H(X,Y) = \log_2 3 - \frac{4}{3}$

(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.

p, q의 확률분포가 다를 때

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

$$D(p\|q) = E_p\left[\log\frac{P(X)}{q(X)}\right] = E_p\left[-\log\frac{q(X)}{P(X)}\right] \geq -\log\left(E_p\left[\frac{q(X)}{P(X)}\right]\right)$$

$$E_p\left[\frac{q(X)}{P(X)}\right] = \sum_x P(x)\frac{q(x)}{P(x)} = \sum_x q(x) = 1 \rightarrow D(p\|q) \geq -\log 1 = 0$$

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