기수 : 15기 이름 : 이윤서

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 $A$에 대하여 다음 조건들은 동치(equivalent)입니다.

(a) 행렬 $A$는 가역적(invertible)이다.

(b) 임의의 $n \times 1$ 벡터 $b$에 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) $A$를 구하고 PD(Positive Definite), PSD, ND, NSD 중 무엇인지 판별하시오.

$\mathrm{A} = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$, $\mathrm{D}_1 = 2 > 0$ , $\mathrm{D}_2 = 3 > 0$ , $\mathrm{D}_3 = \det(\mathrm{A}) > 0$이므로 PD

(2) 위에서 구한 행렬 $A$의 역행렬 $A^{-1}$ 을 계산하시오.

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

(3) 만약 특정 상수 $k$에 대해 함수의 항 $xy$가 $kxy$로 변하여 $\det(A) = 0$이 된다면, 조건 (b)는 여전히 성립하는가? 그 이유를 행렬식(determinant)과 관련지어 설명하시오.

행렬식이 0이면 행렬이 비가역적이므로 역행렬이 존재하지 않는다.

1-2. (Optional)

(1) (a), (b), (c) 세 조건들이 동치임을 증명하시오.

(2) 행렬 $A$가 가역적이라고 가정할 때, 임의의 $b$에 대하여 $Ax = b$를 만족하는 해 $x$가 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 $B^T B, BB^T$ 를 계산하고, 이들의 고유값을 계산해 특이값 $\sigma$와 고유벡터를 통해 직교대각화하여 $B^T B = VDV^T, BB^T = UDU^T$ 를 구합니다.

2) $0$이 아닌 특이값들을 내림차순으로 나열하여 $\Sigma$를 구성하고, 이들을 모두 활용해 $A = U\Sigma V^T$ 를 구합니다.

다음 행렬 $B$에 특이값 분해를 적용하였을 때 나오는 행렬 $U, \Sigma, V^T$ 를 각각 구하시오.

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

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

$$\det\left(\mathrm{B^TB} - \lambda \mathrm{I}\right) = 0$$

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

루트 취하면 $\sigma_1 = \sqrt{2}, \sigma_2 = \sqrt{1} = 1$

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

$$\lambda_1 = 2 \rightarrow v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix},\ \lambda_2 = 1 \rightarrow v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},\ \lambda_3 = 0 \rightarrow v_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$$

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

$$u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

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

# 문제 3    Convex Sets & Functions

## 3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

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

정의에 따라, 집합 내의 임의의 두 점 $x_1, x_2$과 $0 < \theta < 1$ 인 $\theta$에 대해 $x_\theta = \theta x_1 + (1 - \theta)x_2$가 집합에 포함됨을 보이면 된다. $C_1$은 선형부등식으로 정의된 집합이므로 convex 하다. $C_2$는 $|\theta x_1 + (1-\theta)x_2| \leq \theta|x_1| + (1-\theta)|x_2| \leq 1$이 성립하므로 convex 하다. $C_3$는 $g(x) = e^x, g''(x) = e^x > 0$이므로 convex 하다.

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

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

f 가 convex function이므로 $f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta)f(x_2)$
S 의 두 점 $(x_1, y_1), (x_2, y_2)$에 대해 $(x_\theta, y_\theta) = \theta(x_1, y_1) + (1-\theta)(x_2, y_2) = (\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2)$이다. $y_1 \geq f(x_1), y_2 \geq f(x_2)$ 이므로 $y_\theta = \theta y_1 + (1-\theta)y_2 \geq \theta f(x_1) + (1-\theta)f(x_2)$이고, $\theta f(x_1) + (1-\theta)f(x_2) \geq f(\theta x_1 + (1-\theta)x_2) = f(x_\theta)$
따라서 $t_\theta \geq f(x_\theta), (x_\theta, t_\theta) \in S$이므로 집합 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.

$h(\theta x_1 + (1-\theta)x_2) = f(\theta x_1 + (1-\theta)x_2) + g(\theta x_1 + (1-\theta)x_2)$
$\leq \theta f(x_1) + (1-\theta)f(x_2) + \theta g(x_1) + (1-\theta)g(x_2)$
$= \theta(f(x_1) + g(x_1)) + (1-\theta)\big(f(x_2) + g(x_2)\big) = \theta h(x_1) + (1-\theta)h(x_2)$

(b) $A$와 $b$가 호환되는 크기의 행렬과 벡터라고 합시다. $f(x)$가 convex라면 $f(Ax + b)$ 또한 convex 임을 보이시오.

$p(x) = f(Ax + b)$라 하면 $p(\theta x_1 + (1-\theta)x_2) = f(A(\theta x_1 + (1-\theta)x_2) + b) = f(\theta(Ax_1 + b) + (1-\theta)(Ax_2 + b)) \leq \theta f(Ax_1 + b) + (1-\theta)f(Ax_2 + b) = \theta g(x_1) + (1-\theta)g(x_2)$
따라서 f(x)가 convex라면 f(Ax + b) 또한 convex하다.

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.

$f(\theta x_1 + (1-\theta)x_2) = \|A(\theta x_1 + (1-\theta)x_2) - b\|^2$
$= \|\theta(Ax_1 - b) + (1-\theta)(Ax_2 - b)\|^2 \leq \big(\theta\|Ax_1 - b\| + (1-\theta)\|Ax_2 - b\|\big)^2$
$g(u) = u^2 \geq 0$에서 convex하므로 $\big(\theta\|Ax_1 - b\| + (1-\theta)\|Ax_2 - b\|\big)^2 \leq \theta\|Ax_1 - b\|^2 + (1-\theta)\|Ax_2 - b\|^2 = \theta f(x_1) + (1-\theta)f(x_2)$
따라서 f(x)가 convex하다.

기수 : 이름 :

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

$$\mathrm{H(X) = H(Y) = -\left(\frac{2}{3}\log_2\left(\frac{2}{3}\right) + \frac{1}{3}\log_2\left(\frac{1}{3}\right)\right) \approx 0.918}$$

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

$$\mathrm{H(X|Y) = \sum p(y)H(X|Y=y) = \frac{1}{3}\cdot 0 + \frac{2}{3}\cdot 1 = \frac{2}{3} \approx 0.667}$$

$$\mathrm{H(Y|X) = \sum p(x)H(Y|X=x) = \frac{2}{3}\cdot 1 + \frac{1}{3}\cdot 0 = \frac{2}{3} \approx 0.667}$$

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

$$\mathrm{H(X,Y) = -\sum_{x,y} p(x,y)\log_2 p(x,y) = \log_2 3 \approx 1.585}$$

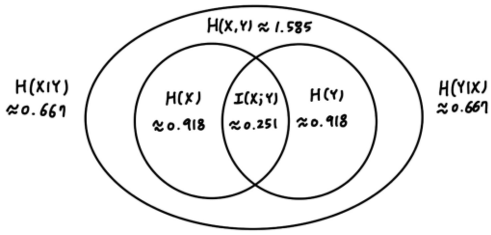
(d) $H(Y) - H(Y \mid X)$.

$$\mathrm{H(Y) - H(Y|X) \approx 0.918 - 0.667 = 0.251}$$

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

$$\mathrm{I(X;Y) = H(Y) - H(Y|X) \approx 0.918 - 0.667 = 0.251}$$

(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.

$\mathrm{p = \{0.2, 0.8\}, q = \{0.6, 0.4\}}$ 일 때, $\mathrm{D(p\,||\,q) = 0.2\log_2\frac{0.2}{0.6} + 0.8\log_2\frac{0.8}{0.4} \approx 0.4830}$ 이고

$\mathrm{D(q\,||\,p) = 0.6\log_2\frac{0.6}{0.2} + 0.4\log_2\frac{0.4}{0.8} \approx 0.5510}$ 이므로 같지 않음

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

$\mathrm{D(p\,||\,q) = \sum_{x\in X} p(x)\log\frac{p(x)}{q(x)}}$, $\mathrm{D(p\,||\,q) = \sum_{x\in X} p(x)\log\frac{p(x)}{q(x)} = -\sum_x p(x)\log\frac{q(x)}{p(x)} = -\sum_x p(x)\log\frac{q(x)}{p(x)}}$

$\mathrm{\geq -\log(\sum_x p(x)\frac{q(x)}{p(x)})}$ , $\mathrm{\sum_x p(x)\frac{q(x)}{p(x)} = \sum q(x) = 1}$

$\mathrm{D(p\,||\,q) \geq -\log(1) = 0}$

따라서 모든 확률분포 p,q 에 대해 $\mathrm{D(p\,||\,q) \geq 0}$ 을 만족 (등호는 p = q 일 때 성립)

## 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