기수 : 14 이름 : 조재우

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.

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

1) $M_1 = 2 > 0$
2) $M_2 = \det\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = 4 - 1 = 3 > 0$
3) $M_3 = \det(A)$
$= 2(6-1) - 1(3-1) + 1(1-2) = 10 - 2 - 1 = 7 > 0$

PD

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

$A^{-1} = \frac{1}{\det(A)} \mathrm{adj}(A) = \frac{1}{7}\begin{bmatrix} 5 & -2 & -1 \\ -2 & 5 & -1 \\ 1 & 1 & 3 \end{bmatrix}$

(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.

성립하지 않음, 1번의 조건에 따르면 행렬 A가 가역적이어야 방정식 Ax=b가 유일한 해를 갖는데 det(A)=0이면 행렬 A는 특이행렬이 되어 역행렬 존재하지 않고 따라서 유일한 해를 갖지 않으므로 조건 b는 성립하지 않음

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 : 14　　　이름 : 조재우

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 구하기 (BB^T의 고유벡터)

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

- 고유값: $\lambda_1 = 2, \lambda_2 = 1$
- 특이값: $\sigma_1 = \sqrt{2}, \sigma_2 = 1$
- 고유벡터: $u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

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

2. Σ 구하기 (특이값)

$\sigma_1 = \sqrt{\lambda_1} = \sqrt{2}$

$\sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1$

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

3. V 구하기 (B^TB의 고유벡터)

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

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

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

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

1) $\lambda_1 = 2$

$\begin{bmatrix} 1-2 & 1 & 0 \\ 1 & 1-2 & 0 \\ 0 & 0 & 1-2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$z = 0$

$-x + y = 0 \Rightarrow x = y$

$\therefore v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

2) $\lambda_2 = 1$

$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$y = 0$, $x = 0$ $\Rightarrow v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

3) $\lambda_3 = 0$

$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$z = 0$

$x + y = 0 \Rightarrow x = -y$

$\therefore v_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}$

$\therefore V = [v_1\ v_2\ v_3] = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}$

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

3-1-1). C_1

임의의 두 점 P1=(x1,y1),P2=(x2,y2)가 C1에 속한다고 가정하자. 즉 x1(2)+y1(2)<=1이고 두점의 내분점 $P_{new} = \theta P_1 + (1-\theta)P_2$ 의 좌표는 $(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2)$ 인데 이 점이 조건을 만족하는지 확인해보자. $(\theta x_1 + (1-\theta)x_2) + (\theta y_1 + (1-\theta)y_2)$ $= \theta(x_1+y_1) + (1-\theta)(x_2+y_2)$ 가정에 의해 x+y<=1이므로: $\leq \theta(1) + (1-\theta)(1) = \theta + 1 - \theta = 1$ 따라서 내분점도 부등식을 만족하므로 Convex Set이다.

C_2

L1 Norm: $||x||_1 = \sum |x_i|$

두 점 $u, v \in C_2$ 라 하면 $||u||_1 \leq 1, ||v||_1 \leq 1$

내분점 $z = \theta u + (1-\theta)v$ 에 대해 삼각 부등식과 노름의 성질을 적용하면 $||\theta u + (1-\theta)v||_1 \leq ||\theta u||_1 + ||(1-\theta)v||_1$

$\theta \geq 0$ 이므로: $= \theta||u||_1 + (1-\theta)||v||_1$ 가정을 대입하면: $\leq \theta(1) + (1-\theta)(1) = 1$ 따라서 Convex Set이다.

C_3

두 점 $(x_1, y_1), (x_2, y_2) \in C_3$ 라 하면 $y_1 \geq e^{x_1}, y_2 \geq e^{x_2}$ 이다.

내분점 $(x_{new}, y_{new}) = (\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2)$ 이다. $y_{new} = \theta y_1 + (1-\theta)y_2 \geq \theta e^{x_1} + (1-\theta)e^{x_2}$ 식이 성립하는지 확인해보면

함수 $f(x) = e^x$ 는 Convex Function이고 $f(x) = e^x > 0$ 정의에 의해 다음이 성립한다. $\theta e^{x_1} + (1-\theta)e^{x_2} \geq e^{\theta x_1 + (1-\theta)x_2}$

따라서 $y_{new} \geq e^{x_{new}}$ 가 성립하므로 Convex Set이다.

3-1-2)

집합 S의 두 원소 (x1,t1),(x2,t2)를 잡으면 조건에 의해 $t_1 \geq f(x_1), t_2 \geq f(x_2)$

내분점 $(x,t) = \theta(x_1,t_1) + (1-\theta)(x_2,t_2)$ 라 할 때, $t \geq f(x)$ 임을 보이면 된다.

f가 Convex function이므로 정의에 의해: $\theta f(x_1) + (1-\theta)f(x_2) \geq f(\theta x_1 + (1-\theta)x_2) = f(x)$

결국 $t \geq f(x)$ 이 성립하므로 Epigraph S는 Convex Set이다.

3-2-1)

a)

h(x) = f(x) + g(x)라 하자. $h(\theta x + (1-\theta)y) = f(\theta x + (1-\theta)y) + g(\theta x + (1-\theta)y)$

$\leq [\theta f(x) + (1-\theta)f(y)] + [\theta g(x) + (1-\theta)g(y)] = \theta[f(x) + g(x)] + (1-\theta)[f(y) + g(y)]$

$= \theta h(x) + (1-\theta)h(y)$

b)

h(x) = f(Ax+b)라 하자. $h(\theta x + (1-\theta)y) = f(A(\theta x + (1-\theta)y) + b)$

행렬 연산의 형성에 의해: $= f(\theta(Ax+b) + (1-\theta)(Ay+b))$

X = Ax+b, Y= Ay+b로 치환 생각하면, f가 Convex하므로: $\leq \theta f(Ax+b) + (1-\theta)f(Ay+b)$

따라서 h(x) Convex입니다.

기수 : 14 이름 : 조재우

문제 3 Convex Sets & Functions

3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

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

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

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

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.

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

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.

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.

P(X=0,Y=0) = 1/3, P(X=0,Y=1) = 1/3, P(X=1,Y=0) = 0, P(X=1,Y=1) = 1/3
P(X=0) = 2/3, P(X=1) = 1/3, P(Y=0) = 1/3, P(Y=1) = 2/3

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

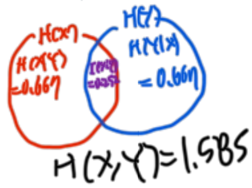
(b) $H(X \mid Y), H(Y \mid X)$. $H(X|Y) = H(X,Y) - H(Y) = \log_2 3 - \left(\log_2 3 - \frac{2}{3}\right) = \mathbf{\frac{2}{3}} \approx 0.667$ bits
$H(Y|X) = H(X,Y) - H(X) = \log_2 3 - \left(\log_2 3 - \frac{2}{3}\right) = \mathbf{\frac{2}{3}} \approx 0.667$ bits

(c) $H(X, Y)$. $H(X,Y) = -\sum\sum p(x,y)\log_2 p(x,y)$ $\quad H(X,Y) = -3 \times (\frac{1}{3}\log_2\frac{1}{3}) = \log_2 3$

(d) $H(Y) - H(Y \mid X)$. $H(Y) - H(Y|X) = \left(\log_2 3 - \frac{2}{3}\right) - \frac{2}{3} = \log_2 3 - \frac{4}{3} \approx 1.585 - 1.333 = \mathbf{0.252}$ bits

(e) $I(X; Y)$. $I(X;Y) = H(X) + H(Y) - H(X,Y)$
$I(X;Y) = \left(\log_2 3 - \frac{2}{3}\right) + \left(\log_2 3 - \frac{2}{3}\right) - \log_2 3 = 2\log_2 3 - \frac{4}{3} - \log_2 3 = \log_2 3 - \frac{4}{3} \approx \mathbf{0.252}$ bits

(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(x) = [0.5,0.5]
q(x) = [0.1,0.9]

$D(p\|q) = \sum p(x)\log\frac{p(x)}{q(x)} = 0.5\log 5 + 0.5\log\frac{5}{9} \approx \mathbf{0.511}$ $\neq$

$D(q\|p) = \sum q(x)\log\frac{q(x)}{p(x)} = 0.1\log\frac{1}{5} + 0.9\log 1.8 \approx \mathbf{0.368}$

$\therefore D(p\|q) \neq D(q\|p)$

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

모르겠습니다...

친구들아 미안해.

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