기수 : 14기 이름 : 고서연

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.

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

(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.

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.

1-1. $f(x,y,z) = x^2+y^2+1.5z^2+xy+yz+zx$

(1)

$$A = \nabla^2 f = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix}$$

$f_x = 2x+y+z$

$f_{xx} = 2$, $f_{xy} = 1$, $f_{xz} = 1$

$f_y = 2y+x+z$

$f_{yy} = 2$, $f_{yz} = 1$, $f_{yx} = 1$

$f_z = 3z+x+y$

$f_{zz} = 3$, $f_{zx} = 1$, $f_{zy} = 1$

$$\therefore A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

$\Delta_1 = 2 > 0$

$\Delta_2 = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 4-1=3 > 0$

$\Delta_3 = \det(A) = 2(6-1)-1(3-1)+1(1-2)$
$= 10-2+-1 = 7 > 0$

$\therefore \Delta_1>0,\ \Delta_2>0,\ \Delta_3>0$ 이므로 A는 PD.

(2) $A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$, $\det(A) = 7$

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

(3) $f_{xy} = k$

$$A(k) = \begin{bmatrix} 2 & k & 1 \\ k & 2 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

$\det(A(k)) = 0$ 이 되는 k $\Rightarrow$ $\det(A(k)) = -(k-2)(3k+4)$

$k=2$ or $-\frac{8}{3}$ 일 때 $\det = 0$

$\therefore \det(A)=0$ 이면 A는 비가역이므로 조건 (b) 성립하지 않음.

기수 : 14기 이름 : 고서연

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.

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

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

$\det(B^TB - \lambda I) = (1-\lambda)\lambda(\lambda-2) = 0$

$\lambda = 2, 1, 0$

i) $\lambda_1 = 2 \rightarrow (B^TB - 2I)v = 0$

$\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ $\quad -v_1 + v_2 = 0$, $v_1 - v_2 = 0$, $-v_3 = 0$

$v = t\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

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

ii) $\lambda_2 = 1 \rightarrow (B^TB - I)v = 0$

$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \quad v_1 = 0, \; v_2 = 0, \; v_3 = 1$

$v = t\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

iii) $\lambda_3 = 0 \rightarrow (B^TB)v = 0$

$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = 0 \quad v_1 + v_2 = 0, \; v_3 = 0$

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

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

$\lambda_1 = 2, \quad \lambda_2 = 1$

i) $\lambda_1 = 2 \rightarrow (BB^T - 2I)u = 0$

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

$\begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0 \quad u_2 = 0. \quad u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

ii) $\lambda_2 = 1 \rightarrow (BB^T - I)u = 0$

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

$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = 0 \quad u_1 = 0 \quad u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

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

$\sigma = \sqrt{\lambda(BB^T)} \Rightarrow \sigma_1 = \sqrt{2}, \; \sigma_2 = 1$

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

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

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

기수 : 14기 이름 : 고서연

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.

3-1. (1)

(a) $C_1 = \{(x,y) \in R^2 : x+y \le 1\}$

임의의 두점 (x_1, y_1), $(x_2, y_2) \in C_1$

$x_1+y_1 \le 1$, $x_2+y_2 \le 1$

$\lambda \in [0,1]$에 대해

$(x,y) = \lambda(x_1,y_1) + (1-\lambda)(x_2,y_2)$

$x = \lambda x_1 + (1-\lambda)x_2$

$y = \lambda y_1 + (1-\lambda)y_2$

$x+y = \lambda(x_1+y_1) + (1-\lambda)(x_2+y_2)$

그런데 $x_1+y_1 \le 1$, $x_2+y_2 \le 1$

$\therefore x+y \le \lambda \cdot 1 + (1-\lambda)\cdot 1 = 1$

$(x,y) \in C_1$

$\therefore C_1$는 convex.

(b) $C_2 = \{x \in R^2 : \|x\|_1 \le 1\}$

$\|x\|_1 = \sum_{i=1}^{n} |x_i|$

임의의 $x_1, x_2 \in C$, 즉 $\|x_1\|_1 \le 1$, $\|x_2\|_1 \le 1$

$x = \lambda x_1 + (1-\lambda)x_2$

$\|x\|_1 = \|\lambda x_1 + (1-\lambda)x_2\|_1$

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

$\le \lambda \cdot 1 + (1-\lambda)\cdot 1 = 1$

$\therefore x \in C_2$

C_2는 convex

(c) $C_3 = \{ (x, y) \in \mathbb{R}^2 : y \ge e^x \}$

e^x는 convex.

임의의 두 점 $(x_1, y_1), (x_2, y_2) \in C_3$

$y_1 \ge e^{x_1}, \quad y_2 \ge e^{x_2}$

$x = \lambda x_1 + (1-\lambda) x_2$

$y = \lambda y_1 + (1-\lambda) y_2$

convex 함수 정의:

$$e^{\lambda x_1 + (1-\lambda) x_2} \le \lambda e^{x_1} + (1-\lambda) e^{x_2}$$

$\therefore y = \lambda y_1 + (1-\lambda) y_2 \ge \lambda e^{x_1} + (1-\lambda) e^{x_2} \ge e^{\lambda x_1 + (1-\lambda) x_2}$

$y \ge e^x$

$\therefore$ C_3는 convex

3-1 (2)

두 점 $(x_1, t_1), (x_2, t_2) \in S$

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

$\lambda \in [0,1]$에 대해 $\quad x = \lambda x_1 + (1-\lambda) x_2$

$t = \lambda t_1 + (1-\lambda) t_2$

convex 함수 정의: $f(x) = f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$

그런데 $t_1 \ge f(x_1), \; t_2 \ge f(x_2)$

$\therefore t = \lambda t_1 + (1-\lambda) t_2 \ge \lambda f(x_1) + (1-\lambda) f(x_2) \ge f(x)$

$\therefore (x, t) \in S$

$\therefore$ S는 convex set

3-2 (1) (a) 가정: f, g는 convex.

임의의 $x_1, x_2, \lambda \in [0,1]$

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

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

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

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

$\therefore f+g$는 convex

(b) $h(x) = f(Ax+b)$

임의의 $x_1, x_2, \lambda \in [0,1]$

$$h(\lambda x_1 + (1-\lambda)x_2) = f(A(\lambda x_1 + (1-\lambda)x_2)+b)$$

$$= f(\lambda Ax_1 + (1-\lambda)Ax_2 + b)$$

$$= f(\lambda(Ax_1+b) + (1-\lambda)(Ax_2+b))$$

$$\le \lambda f(Ax_1+b) + (1-\lambda)f(Ax_2+b)$$

$$= \lambda h(x_1) + (1-\lambda)h(x_2)$$

$\therefore f(Ax+b)$는 convex

3-2 (2)

$$\|y\|^2 = y^T y$$

$$f(x) = (Ax-b)^T(Ax-b)$$

임의의 $x_1, x_2, \lambda \in [0,1]$

$$x = \lambda x_1 + (1-\lambda)x_2$$

$$Ax-b = A(\lambda x_1 + (1-\lambda)x_2) - b$$

$$= \lambda(Ax_1-b) + (1-\lambda)(Ax_2-b)$$

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

벡터 u, v에 대해

$$\| \lambda u + (1-\lambda) v \|^2 = \lambda^2 \|u\|^2 + (1-\lambda)^2 \|v\|^2 + 2\lambda(1-\lambda) u^T v$$

$$u^T v \le \frac{1}{2} (\|u\|^2 + \|v\|^2)$$ 항상 성립.

$$\| \lambda u + (1-\lambda) v \|^2 \le \lambda \|u\|^2 + (1-\lambda) \|v\|^2$$

$u = Ax_1 - b, \quad v = Ax_2 - b$

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

$\therefore f(x) = \|Ax - b\|^2$는 convex.

기수 : 14기 이름 : 고서연

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) $H(X), H(Y)$. $H(X) = -\sum p \log_2 p$

(b) $H(X \mid Y), H(Y \mid X)$. $H(X|Y) = \sum p(y)\, H(X|Y=y)$

(c) $H(X, Y)$. $H(X,Y) = -\sum p(X,Y) \log_2 p(X,Y)$

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

(e) $I(X; Y)$. $I(X;Y) = H(Y) - H(Y|X)$

(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.

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

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

4-1

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

$= -\left(\frac{2}{3}(1-\log_2 3) + \frac{1}{3}(-\log_2 3)\right)$

$= -\left(\frac{2}{3} - \log_2 3\right) = \log_2 3 - \frac{2}{3}$

$H(Y) = \log_2 3 - \frac{2}{3}$

(b)

$Y=0$; $p=\frac{1}{3}$ $\quad P(X=0 \mid Y=0) = \frac{1/3}{1/3} = 1$

$P(X=1 \mid Y=0) = 0$

$Y=1$; $p=\frac{2}{3}$ $\quad P(X=0 \mid Y=1) = \frac{1/3}{2/3} = \frac{1}{2}$

$P(X=1 \mid Y=1) = \frac{1/3}{2/3} = \frac{1}{2}$

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

$X=0$; $p=\frac{2}{3}$ $\quad P(Y=0 \mid X=0) = \frac{1}{2}$

$P(Y=1 \mid X=0) = \frac{1}{2}$

$X=1$; $p=\frac{1}{3}$ $\quad P(Y=0 \mid X=1) = 0$

$P(Y=1 \mid X=1) = 1$

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

(c) $H(X,Y) = -\cancel{3}\cdot\frac{1}{\cancel{3}}\log_2\frac{1}{3} = \log_2 3$

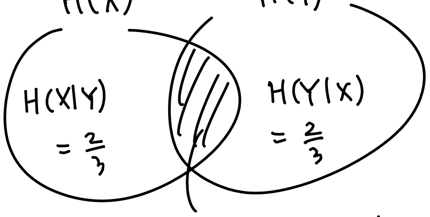
(d) $H(Y) - H(Y \mid X) = \left(\log_2 3 - \frac{2}{3}\right) - \frac{2}{3} = \log_2 3 - \frac{4}{3}$

(e) $I(X;Y) = H(Y) - H(Y \mid X)$

$= \log_2 3 - \frac{4}{3}$

(f)

$H(X) = \log_2 3 - \frac{2}{3}$ $\quad H(Y) = \log_2 3 - \frac{2}{3}$



$I(X;Y) = \log_2 3 - \frac{4}{3}$

4-2 (a)

반례) $p(0)=\frac{1}{2}$, $p(1)=\frac{1}{2}$

$q(0)=\frac{1}{4}$, $q(1)=\frac{3}{4}$

$$D(p\|q)=\frac{1}{2}\log\underbrace{\frac{1/2}{1/4}}_{2}+\frac{1}{2}\log\underbrace{\frac{1/2}{3/4}}_{\frac{2}{3}} \qquad D(q\|p)=\frac{1}{4}\log\underbrace{\frac{1/4}{1/2}}_{\frac{1}{2}}+\frac{3}{4}\log\underbrace{\frac{3/4}{1/2}}_{\frac{3}{2}}$$

$$=\frac{1}{2}\left(\log\frac{4}{3}\right) \neq =\frac{1}{4}\log\frac{1}{2}+\frac{3}{4}\log\frac{3}{2}$$

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

$r(x)=\frac{q(x)}{p(x)}$

$D(p\|q)=-\sum p(x)\log r(x)$

$=-E_p[\log r(x)]$

$E[\log z]\leq \log(E[z])$. log는 concave.

$z=r(x)$

$E_p\left[\log\frac{q(x)}{p(x)}\right]\leq\log\left(E_p\left[\frac{q(x)}{p(x)}\right]\right)$

$-E_p\left[\log\frac{q(x)}{p(x)}\right]\geq-\log\left(\underline{E_p\left[\frac{q(x)}{p(x)}\right]}\right)$

$D(p\|q)\geq$ 〃

$\underline{E_p\left(\frac{q(x)}{p(x)}\right]}=\sum \cancel{p(x)}\frac{q(x)}{\cancel{p(x)}}=1 \Rightarrow -\log(1)=0.$

$\therefore D(p\|q)\geq 0$