기수 : 15 이름 : 박성하.

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) $f_x = 2x + y + z \qquad f_{xx} = 2 \quad f_{xy} = 1 \quad f_{xz} = 1$

$f_y = 2y + x + z \qquad f_{yx} = 1 \quad f_{yy} = 2 \quad f_{yz} = 1$

$f_z = 3z + y + x \qquad f_{zx} = 1 \quad f_{zy} = 1 \quad f_{zz} = 3$

$$\Rightarrow \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

$D_1 = 2 \qquad D_2 = \det\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 4 - 1 = 3$

$D_3 = 2(5) - (3-1) + (1-2) = 10 - 2 - 1 = 7$

$\therefore$ PD

(2) Cofactor matrix.

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

(3) if $\det(A) = 0$, A is singular, A^{-1} doesn't exist.

조건 (b)는 성립X.

1-2 (1) (a)→(b) $Ax = b \Rightarrow x = A^{-1}b$

(b)→(c) $x=0$ 은 $Ax=0$ 을 만족 $x=0$ 는 유일해.

(c)→(a) $Ax=0$ 이 $x=0$을 유일해로 가짐 ⇒ linearly independant.

$\Rightarrow \det(A) \neq 0$.

(2) $Ax = b \Rightarrow A^{-1}(Ax) = A^{-1}b \Rightarrow x = A^{-1}b$

기수 : 15　　이름 : 백성하

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 σ 와 고유벡터를 통해 직교대각화하여 $B^TB = VDV^T, BB^T = UDU^T$ 를 구합니다.
>
> 2) 0 이 아닌 특이값들을 내림차순으로 나열하여 Σ 를 구성하고, 이들을 모두 활용해 $A = U\Sigma V^T$ 를 구합니다.

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

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

$BB' = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow$ eigenvalue $= 2, 1.$

$\lambda_1 = 2 \rightarrow \begin{bmatrix} 1 \\ 0 \end{bmatrix}$　　$\lambda_2 = 1 \rightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix}$　　$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$B'B = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow$ eigenvalue $= 2, 1, 0$

$(1-\lambda)\left((1-\lambda)^2 - 1\right)$　　$2 \rightarrow \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$

$\lambda^2 - 2\lambda$　　$1 \rightarrow \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

$\lambda(\lambda-2)(1-\lambda)$　　$0 \rightarrow \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$

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

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

$U\Sigma V' = B.$

기수 : 15 이름 : 백승하

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

let $p_1 = (x_1, y_1)$ $p_2 = (x_2, y_2) \in C_1$

$[0,1]$의 임의의 θ에 대해 내분점 $p = \theta p_1 + (1-\theta)p_2$ 의 좌표는

$(\theta x_1 + (1-\theta)x_2,\ \theta y_1 + (1-\theta)y_2)$

성분합 $\theta(x_1+y_1) + (1-\theta)(x_2+y_2) \le 1$ $(\because x_1+y_1 \le 1,\ x_2+y_2 \le 1)$

C_1 is convex set.

② $C_2 = \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}$ let $a, b \in C_2$ $\|a\|_1 \le 1$ $\|b\|_1 \le 1$

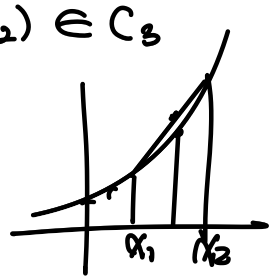
using triangle inequality.

$\|\theta a + (1-\theta)b\|_1 \le \|\theta a\|_1 + \|(1-\theta)b\|_1 \le \theta\|a\|_1 + (1-\theta)\|b\|_1$ $(\because \theta \in [0,1])$

$\le \theta(1) + (1-\theta)(1) = 1$

C_2 is convex set.

③ $C_3 = \{(x,y) \in \mathbb{R}^2 : y \ge e^x\}$. let $(x_1,y_1), (x_2,y_2) \in C_3$



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

$y_{new} = \theta y_1 + (1-\theta)y_2 \ge \theta e^{x_1} + (1-\theta)e^{x_2}$

$\ge e^{\theta x_1 + (1-\theta)x_2} = e^{x_{new}}$.

$\therefore C_3$ is convex set.

2. f : convex. let (x_1, t_1) $(x_2, t_2) \in S$ $t_1 \ge f(x_1)$, $t_2 \ge f(x_2)$ $\theta \in [0,1]$

내분점 (x_{new}, t_{new}) $x_{new} = \theta x_1 + (1-\theta)x_2$ $t_{new} = \theta t_1 + (1-\theta)t_2$

$f(x_{new}) = f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta)f(x_2) \le \theta t_1 + (1-\theta)t_2 = t_{new}$

$\therefore S$ = convex set.

3-2. (a) let $h(x) = f(x)+g(x)$

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

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

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

$\therefore$ $f+g$ is convex set.

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

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

$$= f(\theta(Ax+b)+(1-\theta)(Ax+b))$$

$$\leq \theta \underbrace{f(Ax+b)}_{=h} + (1-\theta)\underbrace{f(Ax+b)}_{=h}$$

$\therefore$ $f(Ax+b)$ is convex.

2. WTS: $g(z)=\|z\|^2$ is convex. $\Rightarrow$ by (b), $\|Ax+b\|^2$ is convex.

let $u = Ax-b$, $v=Ay-b$

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

$$= \theta^2\|u\|^2 + (1-\theta)^2\|v\|^2 + 2\theta(1-\theta)u^Tv \leq \theta\|u\|^2 + (1-\theta)\|v\|^2$$

$$\Rightarrow 0 \leq \theta(1-\theta)\|u\|^2 + \theta(1-\theta)\|v\|^2 - 2\theta(1-\theta)u^Tv$$

$$\Rightarrow 0 \leq \theta(1-\theta)\|u-v\|^2$$

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

기수 : 15 이름 : 박성하 .

문제 4 정보이론 (Information Theory)

4-1. Entropy

확률분포 $p(x, y)$ 가 다음과 같이 주어졌다고 합시다.

$X \backslash Y$	0	1	$P(X)$
0	$\frac{1}{3}$	$\frac{1}{3}$	$2/3$
1	0	$\frac{1}{3}$	$\frac{1}{3}$
$P(Y)$	$\frac{1}{3}$	$\frac{2}{3}$	1

$\rightarrow P(X=0)$

$\rightarrow P(X=1)$

$P(Y=0)$ $P(Y=1)$

$H(X) = -\Sigma\, p(X) \log(p(X))$

$\log 3 - \frac{1}{3}$

다음 값들을 구하시오.

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

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

$H(Y) = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right)$

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

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

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

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

(f) (a)부터 (e)까지 구한 양(quantity)들을 벤 다이어그램으로 나타내시오.

under.

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. (b) $H(X|Y)$ $\quad H(X|Y=0)=0$, $H(X|Y=1)=1$.

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

$H(Y|X)$ $\quad H(Y|X=0)=1$, $H(Y|X=1)=0$.

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

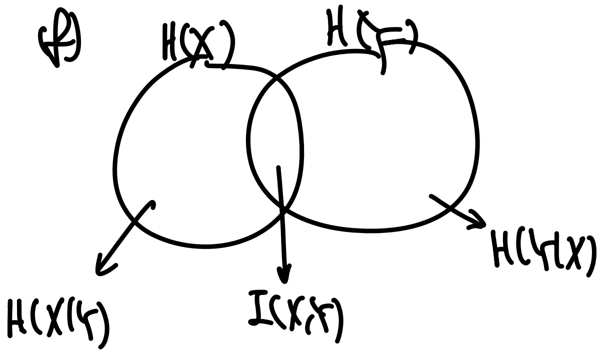
(c) $H(X,Y)= -3\left(\frac{1}{3}\log\frac{1}{3}\right) = -\log\frac{1}{3} = \log 3$

(d) $H(Y) - H(Y|X)$

$$\left(\log 3 - \frac{2}{3}\right) - \frac{2}{3} = \log 3 - \frac{4}{3}.$$

(e) $\log 3 - \frac{4}{3}$.

(f)



4-2. (a) KL-Divergence 는 대칭성이 성립하지 않음.

$$D(p\|q) - D(q\|p) = \Sigma\left((p(x)+q(x))\log p(x) - (p(x)+q(x))\log q(x)\right)$$ ex) if $p(x)=0$, $q(x)>0$.

$\Rightarrow$ not always zero.

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

by Jensen's inequality, $\Sigma p(x)\log\frac{q(x)}{p(x)} \leq \log\left(\Sigma p(x)\frac{q(x)}{p(x)}\right)$

$\Sigma(q(x))=1 \quad \therefore \log(\Sigma q(x)) = 0.$

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