기수 : 14 이름 : 선태희

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.

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

$f_x = 2x+y+z$, $f_y = 2y+x+z$, $f_z = 3z+y+x$

$f_{xx} = 2$, $f_{yy} = 2$, $f_{zz} = 3$

$f_{xy} = f_{yx} = 1$, $f_{xz} = f_{zx} = 1$, $f_{yz} = f_{zy} = 1$

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

$D_1 = 2 > 0$

$D_2 = \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} = 3 > 0$

$D_3 = \det(A) = 2\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} - \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} + \begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} = 7 > 0$

$\therefore$ A is Positive Definite (PD).

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

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

$\det(A(k)) = 2\begin{vmatrix} 2 & 1 \\ 1 & 3 \end{vmatrix} - k\begin{vmatrix} k & 1 \\ 1 & 3 \end{vmatrix} + \begin{vmatrix} k & 2 \\ 1 & 1 \end{vmatrix} = -3k^2+2k+8$

$= -(3k+4)(k-2) = 0 \quad \Leftrightarrow \quad k = 2$ or $-\frac{4}{3}$

If $\det(A(k)) = 0$, $A(k)$ is singular and not invertible.
Hence $A(k)x = b$ does not have a unique solution for all b
thus the condition (b) does not hold.

1-2.

(1) (a) $\Rightarrow$ (b) :

If A is invertible, let $x = A^{-1}b$.
Then, $Ax = A(A^{-1}b) = b$.

Uniqueness: if $Ax = b$ and $Ax' = b$,

$A(x-x') = 0 \Rightarrow x - x' = 0 \Rightarrow x = x'$

(b) $\Rightarrow$ (c) :

Take $b = 0$.
If the solution is unique and $x=0$ is a solution,
it must be the only one.

(c) $\Rightarrow$ (a) :

If $Ax = 0$ has only the trivial solution,
the null space is $\{0\}$ and A is invertible.

$\therefore$ (a), (b), (c) are equivalent.

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.

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

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

$\det(B^TB - \lambda I) = \begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)\begin{vmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{vmatrix} - \begin{vmatrix} 1 & 0 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)^3 - (1-\lambda) = (1-\lambda)\lambda(\lambda-2)$

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

$\sigma_i = \sqrt{\lambda_i} \Rightarrow \sigma_1 = \sqrt{2}, \sigma_2 = 1, \sigma_3 = 0$

(i) $\lambda = 2$: $v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

(ii) $\lambda = 1$: $v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

(iii) $\lambda = 0$: $v_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$

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

$u_i = \frac{1}{\sigma_i} B v_i$

$Bv_1 = \begin{bmatrix} \sqrt{2} \\ 0 \end{bmatrix} \Rightarrow u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $Bv_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Rightarrow u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ $\therefore U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

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

기수 : 14 이름 : 변태희

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.

η-1.

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

Take any $u = (x_1, y_1)$, $v = (x_2, y_2) \in C_1$, $\theta \in [0,1]$.

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

$w = \theta u + (1-\theta)v = (\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2)$.

$w_x + w_y = \theta(x_1+y_1) + (1-\theta)(x_2+y_2) \le \theta \cdot 1 + (1-\theta)\cdot 1 = 1$

So $w \in C_1$ and C_1 is convex.

(ii) $C_2 = \{x \in \mathbb{R}^n \cdot \|x\|_1 \le 1\}$

Take $x_1, x_2 \in C_2$, $\theta \in [0,1]$, let $x = \theta x_1 + (1-\theta)x_2$.

$$\begin{aligned}\|x\|_1 = \|\theta x_1 + (1-\theta)x_2\|_1 &\le \|\theta x_1\|_1 + \|(1-\theta)x_2\|_1 \\ &= \theta\|x_1\|_1 + (1-\theta)\|x_2\|_1 \\ &\le \theta\cdot 1 + (1-\theta)\cdot 1 = 1\end{aligned}$$

Thus $x \in C_2$ and C_2 is convex.

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

Take $u = (x_1, y_1)$, $v = (x_2, y_2) \in C_3$, $\theta \in [0,1]$.

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

$w = \theta u + (1-\theta)v = (\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) = (x_w, y_w)$.

$e^{x_w} = e^{\theta x_1 + (1-\theta)x_2} \le \theta e^{x_1} + (1-\theta)e^{x_2}$ ($\because e^x$ is convex)

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

Therefore, $y_w \ge e^{x_w}$ so $w \in C_3$ and C_3 is convex.

2\. Take any $(x_1, t_1), (x_2, t_2) \in S$, $\theta \in [0,1]$.

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

$(x, t) = \theta(x_1, t_1) + (1-\theta)(x_2, t_2) = (\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2)$.

$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta)f(x_2)$

since $t_i \ge f(x_i)$,

$t = \theta t_1 + (1-\theta)t_2 \ge \theta f(x_1) + (1-\theta)f(x_2) \ge f(\theta x_1 + (1-\theta)x_2)$.

So $(x, t) \in S$ and S is convex.

η-2.

1\. (a) For any x, y and $\theta \in [0,1]$,

$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$,

$g(\theta x + (1-\theta)y) \le \theta g(x) + (1-\theta)g(y)$

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

Thus $f+g$ is convex.

(b) Take any x_1, x_2 and $\theta \in [0,1]$,

$$\begin{aligned}h(\theta x_1 + (1-\theta)x_2) &= f(A(\theta x_1 + (1-\theta)x_2) + b) \\ &= f(\theta(Ax_1+b) + (1-\theta)(Ax_2+b)) \\ &\le \theta f(Ax_1+b) + (1-\theta)f(Ax_2+b) \\ &= \theta h(x_1) + (1-\theta)h(x_2)\end{aligned}$$

$\therefore h$ is convex.

2\. Take any x_1, x_2 and $\theta \in [0,1]$, $u = Ax_1 - b$, $v = Ax_2 - b$

$A(\theta x_1 + (1-\theta)x_2) - b = \theta u + (1-\theta)v$,

$f(\theta x_1 + (1-\theta)x_2) = \|\theta u + (1-\theta)v\|^2$

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

$\theta\|u\|^2 + (1-\theta)\|v\|^2 - \|\theta u + (1-\theta)v\|^2 = \theta(1-\theta)\|u-v\|^2 \ge 0$.

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

$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta)f(x_2)$

$\therefore f(x) = \|Ax - b\|^2$ is convex on $\mathbb{R}^d$.

기수 : 14   이름 : 선태희

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

$H(Y) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3}$

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

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

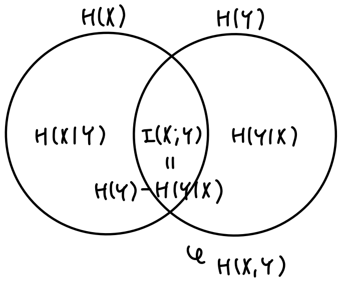
(c) $H(X, Y)$.

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

(d) $H(Y) - H(Y \mid X)$. $= 0.9183 - \frac{2}{3} = 0.2516$

(e) $I(X; Y)$. $= H(X) - H(X|Y) = 0.9183 - \frac{2}{3} = 0.2516$

(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.

eg. $p = (1, 0)$, $q = (\frac{1}{2}, \frac{1}{2})$

$D(p\|q) = \sum_i p_i \log\frac{p_i}{q_i} = 1 \times \log\frac{1}{\frac{1}{2}} = \log 2$ (finite), $D(q\|p) = \sum_i q_i \log\frac{q_i}{p_i}$ contains $q_1 \log\frac{q_1}{p_1} = \frac{1}{2}\log\frac{\frac{1}{2}}{0} = \infty$

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

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

Let $z = \frac{q(X)}{p(X)}$, since $-\log(\cdot)$ is convex, Jensen gives $E_p[-\log z] \geq -\log(E_p[z])$.

$E_p[z] = \sum_x p(x)\frac{q(x)}{p(x)} = \sum_x q(x) = 1$

$\therefore D(p\|q) = E_p[-\log z] \geq -\log 1 = 0$

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