26-1 DSL 정규세션 — Math for ML

- **기수:** 15기
- **이름:** 박현진

문제 1. 선형대수 기초 (Linear Algebra Basic)

1. $A\mathbf{x} = \mathbf{b}$ 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 $\mathbf{b}$에 대하여 방정식 $A\mathbf{x} = \mathbf{b}$는 유일한 해를 갖는다.
- **(c)** 동차방정식(Homogeneous equation) $A\mathbf{x} = \mathbf{0}$은 오직 자명한 해(trivial solution, $\mathbf{x} = \mathbf{0}$)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.

풀이:

편미분을 계산하면 다음과 같다.

$$\begin{aligned} f_x &= 2x + y + z \\ f_y &= x + 2y + z \\ f_z &= x + y + 3z \end{aligned}$$

헤시안 행렬 A는 다음과 같다.

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

주 소행렬식(Leading Principal Minors)을 판별하면:

1. $D_1 = 2 > 0$
2. $D_2 = \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = 4 - 1 = 3 > 0$
3. $D_3 = \det(A) = 2(6-1) - 1(3-1) + 1(1-2) = 10 - 2 - 1 = 7 > 0$

모든 선행 주 소행렬식이 양수이므로, 행렬 A는 **PD (Positive Definite)** 이다.

(2) 위에서 구한 행렬 A의 역행렬 A^{-1}을 계산하시오.

풀이:

$\det(A) = 7$ 이므로, $A^{-1} = \frac{1}{\det(A)}\text{adj}(A)$ 이다. 여인수 C_{ij}를 구하여 전치행렬(수반행렬)을 만들면 된다. A는 대칭행렬이므로 $C_{ij} = C_{ji}$ 이다.

$$
\begin{aligned}
C_{11} &= +(2 \cdot 3 - 1 \cdot 1) = 5 \\
C_{12} &= -(1 \cdot 3 - 1 \cdot 1) = -2 \\
C_{13} &= +(1 \cdot 1 - 2 \cdot 1) = -1 \\
C_{21} &= -(1 \cdot 3 - 1 \cdot 1) = -2 \\
C_{22} &= +(2 \cdot 3 - 1 \cdot 1) = 5 \\
C_{23} &= -(2 \cdot 1 - 1 \cdot 1) = -1 \\
C_{31} &= +(1 \cdot 1 - 2 \cdot 1) = -1 \\
C_{32} &= -(2 \cdot 1 - 1 \cdot 1) = -1 \\
C_{33} &= +(2 \cdot 2 - 1 \cdot 1) = 3
\end{aligned}
$$

따라서 역행렬은 다음과 같다.

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

(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.

풀이:

행렬식 $\det(A) = 0$ 이라는 것은 행렬 A의 역행렬(A^{-1})이 존재하지 않음을 의미한다.

조건 (b)는 임의의 벡터 $\mathbf{b}$에 대하여 방정식 $A\mathbf{x} = \mathbf{b}$가 유일한 해를 가져야 한다고 명시하는데, 유일한 해 $\mathbf{x}$를 가지려면 $\mathbf{x} = A^{-1}\mathbf{b}$ 형태로 유일하게 도출할 수 있어야 한다.

하지만 $\det(A) = 0$ 이 되어 역행렬이 존재하지 않으므로, 유일한 해를 구할 수 없다. 따라서 조건 (b)는 성립하지 않는다.

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.

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

풀이:

1. U

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

BB^T는 대각행렬이므로 고유값과 고유벡터는 다음과 같다.

- $\lambda_1 = 2$ 일 때, 고유벡터 $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$
- $\lambda_2 = 1$ 일 때, 고유벡터 $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$

따라서 U는 다음과 같다.

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

2. Σ

특이값은 고유값의 제곱근이므로:

- $\sigma_1 = \sqrt{\lambda_1} = \sqrt{2}$
- $\sigma_2 = \sqrt{\lambda_2} = \sqrt{1} = 1$

B가 2×3 행렬이므로 Σ도 2×3 행렬이다.

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

문제 3. Convex Sets & Functions

3-1. Convex Set

1. 다음 집합들이 Convex set인지 증명하시오.

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

풀이:

두 점 $(x_1, y_1), (x_2, y_2) \in C_1$ 이라 하면, $x_1 + y_1 \le 1$ 이고 $x_2 + y_2 \le 1$ 이다.

내분점 $(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2)$ 에 대해 식에 대입하면:

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

부등식을 만족하므로 C_1은 Convex Set이다.

$$C_2 = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \le 1\}$$

풀이:

두 벡터 $u, v \in C_2$ 이라 하면, $\|u\|_1 \le 1$ 이고 $\|v\|_1 \le 1$ 이다.
내분점 $\theta u + (1-\theta)v$ 의 Norm에 삼각부등식을 적용하면:

$$\|\theta u + (1-\theta)v\|_1 \le \|\theta u\|_1 + \|(1-\theta)v\|_1 = \theta\|u\|_1 + (1-\theta)\|v\|_1 \le \theta(1) + (1-\theta)(1) = 1$$

부등식을 만족하므로 C_2는 Convex Set이다.

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

풀이:

두 점 $(x_1, y_1), (x_2, y_2) \in C_3$ 이라 하면, $y_1 \ge e^{x_1}$ 이고 $y_2 \ge e^{x_2}$ 이다.
지수 함수 e^x 자체가 Convex 함수이므로, 내분점의 x 좌표에 대해 다음이 성립한다.

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

이때 y_1, y_2의 조건에 의해 $\theta e^{x_1} + (1-\theta)e^{x_2} \le \theta y_1 + (1-\theta)y_2$ 이다.
따라서 $\theta y_1 + (1-\theta)y_2 \ge e^{\theta x_1 + (1-\theta)x_2}$ 가 되어 내분점도 C_3에 속하므로 Convex Set이다.

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

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

풀이:

두 점 $(x_1, t_1), (x_2, t_2) \in S$ 이라 하면, 정의에 의해 $t_1 \geq f(x_1)$ 이고 $t_2 \geq f(x_2)$ 이다.
내분점 $(\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2)$ 가 S에 속하는지 확인해야 한다.

f가 Convex 함수이므로:

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

여기에 초기 조건을 대입하면:

$$\theta f(x_1) + (1-\theta)f(x_2) \leq \theta t_1 + (1-\theta)t_2$$

결과적으로 $f(\theta x_1 + (1-\theta)x_2) \leq \theta t_1 + (1-\theta)t_2$ 이 성립하므로, 집합 S는 Convex Set이다.

3-2. Convex Function

1. Convex 함수의 합성

$f : \mathbb{R}^d \to \mathbb{R}$가 실수값을 갖는 함수라고 합시다.

(a) 함수 $f(x)$와 $g(x)$가 convex일 때, $f(x) + g(x)$ 또한 convex임을 보이시오.

풀이:

$h(x) = f(x) + g(x)$ 라고 하자.

$$\begin{aligned} h(\theta x + (1-\theta)y) &= f(\theta x + (1-\theta)y) + g(\theta x + (1-\theta)y) \\ &\leq (\theta f(x) + (1-\theta)f(y)) + (\theta g(x) + (1-\theta)g(y)) \quad (\because f, g \text{ are convex}) \\ &= \theta(f(x) + g(x)) + (1-\theta)(f(y) + g(y)) \\ &= \theta h(x) + (1-\theta)h(y) \end{aligned}$$

따라서 $f(x) + g(x)$는 Convex이다.

(b) A와 $\mathbf{b}$가 호환되는 크기의 행렬과 벡터라고 합시다. $f(\mathbf{x})$가 convex라면 $f(A\mathbf{x} + \mathbf{b})$ 또한 convex임을 보이시오.

풀이:

$h(x) = f(Ax + b)$ 라고 하자.

$$\begin{aligned} h(\theta x + (1-\theta)y) &= f(A(\theta x + (1-\theta)y) + b) \\ &= f(\theta(Ax + b) + (1-\theta)(Ay + b)) \\ &\leq \theta f(Ax + b) + (1-\theta)f(Ay + b) \quad (\because f \text{ is convex}) \\ &= \theta h(x) + (1-\theta)h(y) \end{aligned}$$

따라서 $f(Ax + b)$는 Convex이다.

2. Convex Optimization

또 다른 함수 $f : \mathbb{R}^d \to \mathbb{R}$를 고려해 봅시다.

$$f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2$$

오직 convex 함수의 정의만을 사용하여, $f(\mathbf{x})$가 $\mathbf{x} \in \mathbb{R}^d$에서 convex임을 보이시오.

풀이:

정의에 따라 $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$ 임을 보여야 한다.
$z_1 = Ax - b$, $z_2 = Ay - b$ 로 치환하자.

우변(RHS) - 좌변(LHS)을 계산하여 0 이상임을 증명하겠다.

$$\begin{aligned}
\mathrm{I} &= \theta\|z_1\|^2 + (1-\theta)\|z_2\|^2 - \|\theta z_1 + (1-\theta)z_2\|^2 \\
&= \theta\|z_1\|^2 + (1-\theta)\|z_2\|^2 - (\theta^2\|z_1\|^2 + (1-\theta)^2\|z_2\|^2 + 2\theta(1-\theta)\langle z_1, z_2\rangle) \\
&= (\theta - \theta^2)\|z_1\|^2 + ((1-\theta) - (1-\theta)^2)\|z_2\|^2 - 2\theta(1-\theta)\langle z_1, z_2\rangle \\
&= \theta(1-\theta)\|z_1\|^2 + (1-\theta)\theta\|z_2\|^2 - 2\theta(1-\theta)\langle z_1, z_2\rangle \\
&= \theta(1-\theta)(\|z_1\|^2 + \|z_2\|^2 - 2\langle z_1, z_2\rangle) \\
&= \theta(1-\theta)\|z_1 - z_2\|^2
\end{aligned}$$

이때 $\theta \in [0,1]$ 이므로 $\theta(1-\theta) \geq 0$ 이고, 노름의 제곱인 $\|z_1 - z_2\|^2 \geq 0$ 이다.
따라서 우변 - 좌변 ≥ 0 이 되므로, $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$ 가 성립하여 $f(x)$는 Convex 이다.

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) $H(X)$, $H(Y)$

풀이:

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

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

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

풀이:

조건부 확률에 대한 엔트로피를 계산한 후, 주변 확률을 곱하여 구한다.

$Y = 0$일 때 X는 무조건 0이므로 불확실성이 없다. $H(X \mid Y = 0) = 0$
$Y = 1$일 때 X는 0일 확률 $\frac{1}{2}$, 1일 확률 $\frac{1}{2}$이므로 엔트로피는 1이다. $H(X \mid Y = 1) = 1$

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

$X = 0$일 때 Y는 0일 확률 $\frac{1}{2}$, 1일 확률 $\frac{1}{2}$이다. $H(Y \mid X = 0) = 1$
$X = 1$일 때 Y는 무조건 1이다. $H(Y \mid X = 1) = 0$

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

(c) $H(X, Y)$

풀이:

결합 확률 분포에서 0이 아닌 확률은 모두 $\frac{1}{3}$이며 총 3개다.

$$H(X, Y) = -\sum_{x,y} p(x, y)\log_2 p(x, y) = -3 \times \left(\frac{1}{3}\log_2\frac{1}{3}\right) = -\log_2\frac{1}{3} = \log_2 3$$

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

풀이:

(a)와 (b)에서 구한 값을 대입한다.

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

(e) $I(X; Y)$

풀이:

상호정보량(Mutual Information)의 정의에 의해 $I(X;Y) = H(Y) - H(Y \mid X)$ 와 동일하다.

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

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

풀이:

4-2. KL-divergence

(a) $D(q\|p) = D(p\|q)$가 성립하지 않는 반례를 제시하시오.

풀이:

$D(q\|p) \neq D(p\|q)$ 인 반례를 보이겠다.
간단한 베르누이 분포 $p = (0.5, 0.5)$, $q = (0.9, 0.1)$ 라고 가정하자.

$$D(p\|q) = 0.5\log\left(\frac{0.5}{0.9}\right) + 0.5\log\left(\frac{0.5}{0.1}\right) \approx 0.73$$
$$D(q\|p) = 0.9\log\left(\frac{0.9}{0.5}\right) + 0.1\log\left(\frac{0.1}{0.5}\right) \approx 0.53$$

두 값이 다르므로 대칭성이 성립하지 않는다.

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

풀이:

함수 $f(t) = -\log(t)$는 strictly convex 함수이다.

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

기댓값 형태로 표현하면 $E_p\left[-\log\frac{q(x)}{p(x)}\right]$ 이다.
Jensen's Inequality에 의해:

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

여기서 괄호 안의 기댓값을 계산하면:

$$E_p\left[\frac{q(x)}{p(x)}\right] = \sum p(x)\frac{q(x)}{p(x)} = \sum q(x) = 1$$

따라서 다음이 성립한다.

$$D(p\|q) \geq -\log(1) = 0$$