Math for ML

구기현

문제 1 선형대수 기초 (Linear Algebra Basic)

1-1

(1)

$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 3 \end{pmatrix}$ 이고,

$D_1 = \det(2) = 2 > 0$,

$D_2 = \det\left(\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right) = (2 \times 2) - (1 \times 1) = 3 > 0$,

$D_3 = \det(A) > 0$

이므로, 정의에 따라 이 행렬은 PD 이다

(2)

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

(3)

Determinant 가 0 이면 성질에 따라 역행렬이 존재하지 않는다.

문제 2 특이값 분해 (SVD)

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

$\det(B^T B - \lambda I) = 0$ 을 풀면 $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 0$ 이다.

따라서 $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sqrt{1} = 1$ 이다.

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

$\lambda_1 = 2$ 일 때, $v_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$

$\lambda_2 = 1$ 일 때, $v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$

$\lambda_3 = 0$ 일 때, $v_3 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}$

따라서 $V = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}$ 이며, $V^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$ 이다.

$u_i = \frac{1}{\sigma_i} B v_i$ 이므로, $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ 이다. 따라서 $U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

결과적으로

$$B = U \Sigma V^T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

문제 3 Convex Sets & Functions

3-1. Convex Set

1. Convex set의 정의에 따라, 집합 내의 임의의 두 점 x_1, x_2와 $0 < \theta < 1$ 인 θ에 대해 $x_\theta = \theta x_1 + (1-\theta)x_2$가 다시 집합에 포함됨을 보이면 됨.

C_1 은 Half-space이다. 따라서 convex 함

C_2 도 convex 함. 삼각부등식에 의해 $|\theta x_1 + (1-\theta)x_2| \le \theta|x_1| + (1-\theta)|x_2| \le 1$ 이 성립.

C_3 도 convex 함. $g(x) = e^x$, $g''(x) = e^x > 0$ 이므로

2. f 는 convex function이므로, $f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta)f(x_2)$ 가 성립.

S 의 두 점 $(x_1, t_1), (x_2, t_2)$를 잡으면 $(x_\theta, t_\theta) = \theta(x_1, t_1) + (1-\theta)(x_2, t_2) = (\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2)$이다.

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

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

$\theta f(x_1) + (1-\theta)f(x_2) \ge f(\theta x_1 + (1-\theta)x_2) = f(x_\theta)$

따라서 $t_\theta \ge f(x_\theta)$ $therefore$ $(x_\theta, t_\theta) \in S$

3-2. Convex Function

1. Convex 함수의 합성

(a) $h(\theta x_1 + (1-\theta)x_2) = f(\theta x_1 + (1-\theta)x_2) + g(\theta x_1 + (1-\theta)x_2)$

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

$= \theta(f(x_1) + g(x_1)) + (1-\theta)\big(f(x_2) + g(x_2)\big) = \theta h(x_1) + (1-\theta)h(x_2)$

(b) $g(x) = f(Ax + b)$ 라 가정

$g(\theta x_1 + (1-\theta)x_2) = f(A(\theta x_1 + (1-\theta)x_2) + b) = f(\theta(Ax_1 + b) + (1-\theta)(Ax_2 + b))$

$\leq \theta f(Ax_1 + b) + (1-\theta)f(Ax_2 + b)$

$= \theta g(x_1) + (1-\theta)g(x_2)$

따라서 convex

2. Convex Optimization

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &= \left|\left|A(\theta x_1 + (1-\theta)x_2) - b\right|\right|^2 \\ &= \left|\left|\theta(Ax_1 - b) + (1-\theta)(Ax_2 - b)\right|\right|^2 \\ &\leq \left(\theta \left|\left|Ax_1 - b\right|\right| + (1-\theta)\left|\left|Ax_2 - b\right|\right|\right)^2 \end{aligned}$$

이때 $h(u) = u^2 \geq 0$ 에서 convex 한 함수이므로

$$\begin{aligned} \left(\theta \left|\left|Ax_1 - b\right|\right| + (1-\theta)\left|\left|Ax_2 - b\right|\right|\right)^2 &\leq \theta \left|\left|Ax_1 - b\right|\right|^2 + (1-\theta)\left|\left|Ax_2 - b\right|\right|^2 \\ &= \theta f(x_1) + (1-\theta)f(x_2) \end{aligned}$$

따라서 $f(x)$ 는 convex

문제 4 정보이론 (Information Theory)

4-1. Entropy

(a) $H(X) = H(Y) = -\left(\frac{2}{3} \, log_2\left(\frac{2}{3}\right) + \frac{1}{3} \, log_2\left(\frac{1}{3}\right)\right) \approx 0.918$

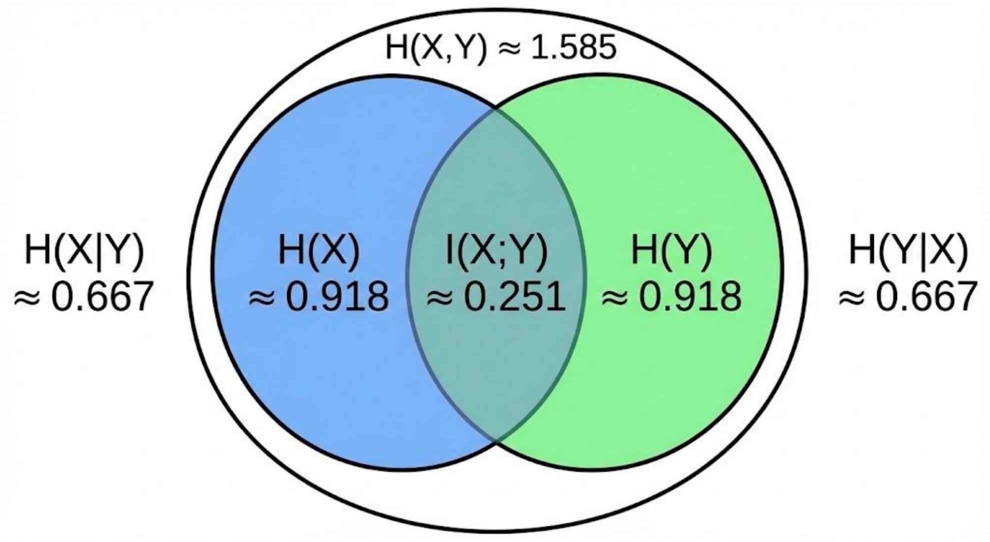
(b) $H(Y|X) = \sum p(x)H(Y|X=x) = \frac{2}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{2}{3} \approx 0.667$

(c) $H(X,Y) = -\sum_{x,y} p(x,y) log_2 \, p(x,y) = log_2 \, 3 \approx 1.585$

(d) $H(Y) - H(Y|X) \approx 0.918 - 0.667 = 0.251$

(e) $I(X;Y) = H(Y) - H(Y|X) \approx 0.918 - 0.667 = 0.251$

(f)



4-2. KL-divergence

(a) 만약 $p = \{0.1, 0.9\}$, $q = \{0.5,\ 0.5\ \}$ 라면,

$D(p\,||\,q) = \sum_{x\in X} p(x)\, log_2 \frac{p(x)}{q(x)}$ 이므로,

$$D(p\,||\,q) = 0.1 \log_2 \frac{0.1}{0.5} + 0.9 \log_2 \frac{0.9}{0.5} \approx 0.531$$

$$D(q\,||\,p) = 0.5 \log_2 \frac{0.5}{0.1} + 0.5 \log_2 \frac{0.5}{0.9} \approx 0.737$$

로 둘은 같지 않음

(b) $D(p\,||\,q) = \sum_{x\in X} p(x)\, log \frac{p(x)}{q(x)}$

$D(p\,||\,q) = \sum_{x\in X} p(x)\, log \frac{p(x)}{q(x)} = -\sum_x p(x)\, log \frac{q(x)}{p(x)} = -\sum_x p(x) log \frac{q(x)}{p(x)} \geq -\log\,(\sum_x p(x) \frac{q(x)}{p(x)})$

$$\sum_x p(x) \frac{q(x)}{p(x)} = \sum q(x) = 1$$

따라서 $D(p\,||\,q) \geq -\log(1) = 0$

모든 확률분포 p,q 에 대해 $D(p\,||\,q) \geq 0$ 을 만족하고, 등호는 p = q 일 때 성립.