Homework 1

Park Jae Un
2019-12551

4190.408 001, Artificial Intelligence
September 15, 2025

1 Linear algebra

1. **(Matrix norm) Show that for matrix $A \in \mathbb{R}^{n \times n}$ and its maximum singular value $\sigma_{\max}$, the following holds:**

$$\|A\|_2 = \sigma_{\max}.$$

spectral norm은

$$\|A\|_2 = \sup_{\|x\|=1} \|Ax\|_2$$

으로 정의된다. $A = U\Sigma V^T$ (SVD)라 두고 $y = V^T x$라 하면

$$\|Ax\|_2 = \|\Sigma y\|_2, \quad \|\Sigma y\|_2^2 = \sum_i \sigma_i^2 y_i^2.$$

y의 길이가 1일 때 이 값의 최대는 가장 큰 singular value $\sigma_{\max}$에서 달성된다. 따라서

$$\|A\|_2 = \sigma_{\max}.$$

2. $(Ax = 0)$ **For some $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, solve**

$$\arg\min_{\|x\|=1} \|Ax\|_2^2$$

using SVD.

$A = U\Sigma V^T$, $x = Vy$, $\|y\| = 1$이라 하면

$$\|Ax\|_2^2 = \sum_i \sigma_i^2 y_i^2.$$

최소값은 $\sigma_{\min}$에 대응하는 singular vector 방향에서 얻어진다. 만약 $\sigma_{\min} = 0$이면 해는 $\ker(A)$에 속하는 단위 벡터들이다.

3. $(Ax = b)$ **Derive the pseudo-inverse of $A = U\Sigma V^T$.**

A의 pseudo-inverse는

$$A^+ = V\Sigma^+ U^T,$$

여기서

$$\Sigma^+ = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)^T.$$

이를 사용하면 minimum least-squares 해이자 minimal norm 해는

$$x^* = A^+ b.$$

2 Probability

1. **(Bayes' theorem) A factory has M_1, M_2, M_3 with priors $0.2, 0.3, 0.5$, and failure rates $0.03, 0.02, 0.01$. Find $P(M_i|F)$ when a failed product is observed.**

전체 불량 확률은

$$P(F) = 0.2 \cdot 0.03 + 0.3 \cdot 0.02 + 0.5 \cdot 0.01 = 0.017.$$

Bayes' theorem에 의해

$$P(M_1|F) = 6/17, \quad P(M_2|F) = 6/17, \quad P(M_3|F) = 5/17.$$

2. **(Gaussian distributions) If $X, Y \sim N(0,1)$ independent, show $Z = X + Y$ is Gaussian.**

X, Y가 independent $N(0,1)$이라면

$$M_X(t) = e^{t^2/2}, \quad M_Y(t) = e^{t^2/2}.$$

따라서

$$M_Z(t) = M_X(t)M_Y(t) = e^{t^2}.$$

이는 $N(0,2)$의 MGF와 같으므로

$$Z \sim N(0,2).$$

3. **(KL Divergence) Show that for $p, q \in \mathbb{R}^n$, $D_{KL}(p\|q) \geq 0$.**

정의:

$$D_{KL}(p\|q) = \sum_i p_i \log \frac{p_i}{q_i}.$$

$-\log$가 convex function이므로 Jensen's inequality를 적용하면

$$D_{KL}(p\|q) \geq 0.$$

등호는 $p = q$일 때만 성립한다.

3 Optimization

1. **(Directional derivative and steepest descent) Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. (a) Show $D_v f(x) = \nabla f(x)^T v$. (b) Among all unit vectors u with $\|u\|_2 = 1$, show that the steepest descent direction at x is $-\nabla f(x)/\|\nabla f(x)\|_2$.**

   (a) $g(t) = f(x + tv)$라 하면

   $$D_v f(x) = g'(0) = \nabla f(x)^T v.$$

   (b) $\|u\| = 1$일 때

   $$\nabla f(x)^T u \geq -\|\nabla f(x)\|_2.$$

   최솟값은

   $$u^* = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}.$$

2. **(Convex minimizer) Show that any local minimizer of a convex function is also global.**

   $f(y) < f(x^*)$라 가정하면

   $$f(tx^* + (1-t)y) \leq tf(x^*) + (1-t)f(y) < f(x^*).$$

   이는 모순이므로 x^*는 global minimizer이다.