Machine Learning
Homework 1

Tunyu Zhang PB21081566

September 26, 2023

1. Exercise 1: Bolzano-Weierstrass Theorem

1.1 Solution

We assume that

$$u \neq sup\mathbf{C}$$

This means $\exists u' \in \mathbb{R}$, u' is an upper bound of set $\mathbf{C}$ and $u' \leq u$. We have already known that: u is an upper bound of set $\mathbf{C}$ and

$$\forall \epsilon > 0, \exists a \in \mathbf{C}, \quad \text{such that} \quad u - a < \epsilon \tag{1.1}$$

Now we let $\epsilon_0 = u - u' > 0$ (Because $u' < u$).According to 1.1 we know that:

$$\exists a \in \mathbf{C}, \quad \text{such that} \quad u - a < \epsilon_0 = u - u'$$

It means that:

$$u' < a$$

This is inconsistent with the premise that u is an upper bound.

1.2 Solution

Let $a_1 = a$ and $b_1 = b$. We divide interval $[a_1, b_1]$ into 2 intervals: $[a_1, \frac{a_1+b_1}{2}]$ and $[\frac{a_1+b_1}{2}, b_1]$. After that we let $a_2 = a_1$ and $b_2 = \frac{a_1+b_1}{2}$, so we get the second interval: $[a_2, b_2]$. And it is obvious that $[a_2, b_2] \subset [a_1, b_1]$. Then we divide the second interval $[a_2, \frac{a_2+b_2}{2}]$ and $[\frac{a_2+b_2}{2}, b_2]$, and we let $a_3 = a_2$ and $b_3 = \frac{a_2+b_2}{2}$ to get the third interval $[a_3, b_3]$. We keep doing this operation and obtain a sequence of intervals $\{[a_n, b_n]\}$, and:

$$......[a_n, b_n] \subset [a_{n-1}, b_{n-1}]......[a_2, b_2] \subset [a_1, b_1] \tag{1.2}$$

Besides, $b_n - a_n = \frac{b_{n-1}-a_{n-1}}{2} = \frac{b_{n-2}-a_{n-2}}{2^2} = ... = \frac{b_1-a_1}{2^{n-1}}$, it means:

$$\lim_{n\to\infty} b_n - a_n = 0 \tag{1.3}$$

According to 1.2 and 1.3, we know the sequence $\{[a_n, b_n]\}$ is a closed nested interval.The principle of nested intervals tells us that:

$$\exists \xi \in [a_n, b_n], n = 1, 2, 3...., \quad \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = \xi \tag{1.4}$$

In every interval $[a_k, b_k]$, we choose a element of sequence (x_n) as x_{k_1}. Sequence (x_{n_k}) is a subsequence of (x_n). Note that $a_k \leq x_k \leq b_k$. According to 1.4 and **Squeeze Theory**, we know:

$$\exists c \in [a, b] \quad \lim_{k\to\infty} x_{n_k} = c$$

2. Exercise 2: Limit and Limit Points

2.1 Solution

Lemma 1. A bounded sequence has at least one gathering point

Proof: **Exercise 1: 2** tells us that a bounded sequence (x_n) must have a sequence that converges to c. According to the definition of Limit point, we know c is the limit point of (x_n).

Sufficiency is obvious, the following proves the necessity.

Assume that sequence $\{\mathbf{x}_n\}$ doesn't converge to $\mathbf{x}$. For $\epsilon_0 > 0$, there are infinite points outside the ϵ–neighborhood of limit point $\mathbf{x}$, denoted $\{\mathbf{x}_{n_k}\}$. Obviously, sequence $\{\mathbf{x}_{n_k}\}$ is also a bounded sequence, so it must have a limit point $\mathbf{y}$. And $\mathbf{y} \notin N_\epsilon(\mathbf{x})$. It means $\{\mathbf{x}_n\}$ has 2 different limit points, which is inconsistent with the prerequisite.

2.2 Solution

(1) It is obvious that $\forall x \in [0, 1] \quad \forall \epsilon > 0 \quad N_\epsilon(x) \cap \mathbf{C} \neq \emptyset$, and $N_\epsilon(x) \cap (0, 1) \neq \{x\}$

For $x = 2$, $\exists 0 < \epsilon_0 < 1$, $N_\epsilon(2) \cap \mathbf{C} = 2$, so $x = 2$ is not limit point.

To sum up, $\mathbf{C}' = [0, 1]$ and $x = 2$ is isolated point of $\mathbf{C}'$.

(2) Let the set of limit points of C' be $(C')'$. According to the definition of limit points, we have:

$$\forall p_0 \in (C')', \forall \epsilon > 0 \quad \exists p_1 \in C' \quad \text{such that} \quad p_1 \in N_\epsilon(p_0)$$

Because p_1 is also the limit point of C, we have:

$$\exists p_2 \in C \quad \text{such that} \quad p_2 \in N_{\epsilon - d(p_0, p_1)}(p_1)$$

It means $p_2 \in N_\epsilon(p_0)$. In other words, p_0 is the limit of C. Or we can say $p_0 \in C'$. To sum up, C' contains all of its limit points. C' is a closed set.

3. Exercise 3: Norms

3.1 Solution

(a) It is obvious that $|x_i|^p \geq 0, \quad n = 1,2,3.......$ So $\|\mathbf{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} \geq 0$, and $\|\mathbf{x}\|_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} = 0$ if and only if $\mathbf{x} = 0$. (Nonnegative and definite)

Next we show the homogeneous:

$$\forall \mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R} \qquad \|t\mathbf{x}\|_p = (\sum_{i=1}^{n} |tx_i|^p)^{\frac{1}{p}} = |t|(\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} = |t|\|\mathbf{x}\|_p$$

Finally we show it satisfies the triangle inequality. Let's set $\mathbf{x} = (x_1, x_2, x_3, ..., x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, y_2, y_3, ..., y_n) \in \mathbb{R}^n$.

When $\|\mathbf{x}+\mathbf{y}\|_p = 0$, according to non-negativity, it is trivial. When $\mathbf{x} = 0$ or $\mathbf{y} = 0$, the inequality holds obviously. So we just think about those cases that $\|\mathbf{x}+\mathbf{y}\|_p \neq 0$ and $\mathbf{x} \neq 0$ and $\mathbf{y} \neq 0$.

We know that the absolute value function satisfies the triangle inequality:

$$\begin{aligned} |x_i + y_i| &\leq |x_i| + |y_i| \quad i = 1,2,3,...n \\ |x_i + y_i|^p &\leq |x_i|^p + |y_i|^p \end{aligned}$$

So we have:

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|_p^p &= \sum_{i=1}^{n} |x_i + y_i||x_i + y_i|^{p-1} \\ &\leq \sum_{i=1}^{n} |x_i||x_i + y_i|^{p-1} + \sum_{i=1}^{n} |y_i||x_i + y_i|^{p-1} \\ &= \sum_{i=1}^{n} |x_i(x_i + y_i)^{p-1}| + \sum_{i=1}^{n} |y_i(x_i + y_i)^{p-1}| \quad (|z_1 z_2| = |z_1||z_2|) \\ &= \|\mathbf{x}^{\mathbf{T}}(\mathbf{x}+\mathbf{y})^{\mathbf{p-1}}\|_1 + \|\mathbf{y}^{\mathbf{T}}(\mathbf{x}+\mathbf{y})^{\mathbf{p-1}}\|_1 \end{aligned}$$

We let $q \in \mathbb{R}_{>0}$ which satisfies: $\frac{1}{p} + \frac{1}{q} = 1$. From Hölder's Inequality for Sums we know $\|\mathbf{xy}\|_1 \leq \|\mathbf{x}\|_p\|\mathbf{y}\|_q$. Thus we can derive:

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|_p^p &= \|\mathbf{x}^{\mathbf{T}}(\mathbf{x}+\mathbf{y})^{\mathbf{p-1}}\|_1 + \|\mathbf{y}^{\mathbf{T}}(\mathbf{x}+\mathbf{y})^{\mathbf{p-1}}\|_1 \\ &\leq \|\mathbf{x}\|_p\|(\mathbf{x}+\mathbf{y})^{p-1}\|_q + \|\mathbf{y}\|_p\|(\mathbf{x}+\mathbf{y})^{p-1}\|_q \end{aligned}$$

Note that $q(p-1) = 1$, so $\|(\mathbf{x}+\mathbf{y})^{p-1}\|_q = \|\mathbf{x}\|_p\|\mathbf{x}+\mathbf{y}\|_p^{p-1}$. It means:

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|_p^p &\leq \|\mathbf{x}\|_p\|\mathbf{x}+\mathbf{y}\|_p^{p-1} + \|\mathbf{y}\|_p\|\mathbf{x}+\mathbf{y}\|_p^{p-1} \\ \|\mathbf{x}+\mathbf{y}\|_p &\leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p \end{aligned}$$

(b) We let $M = max_{1\le i\le n}|x_i|$. $\|\mathbf{x}\|_p = (\sum_{i=1}^{n}|x_i|^p)^{\frac{1}{p}} = M + N$ and obviously $N \ge 0$. Thus we know $M \le \|\mathbf{x}\|_p$. So we can say $M \le \lim_{p\to\infty}\|\mathbf{x}\|_p$.

On the other hand,

$$\begin{aligned}\lim_{p\to\infty}\|\mathbf{x}\|_p &= M\lim_{p\to\infty}(\sum_{i=1}^{n}|\frac{|x_i|}{M}|^p)^{\frac{1}{p}}\\ &\le M\lim_{p\to\infty}(\sum_{i=1}^{n}|\frac{M}{M}|^p)^{\frac{1}{p}}\\ &= M\end{aligned}$$

According to **Squeeze Theory**, we can derive $\lim_{p\to\infty}\|\mathbf{x}\|_p = M$.

3.2 Solution

(a) Let's set:

$$\begin{aligned}\mathbf{x} &= (x_1, x_2, ..., x_n)^T\\ \mathbf{A} &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \vdots & \vdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}\\ \mathbf{Ax} &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\\ \vdots\\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n\end{pmatrix} = \begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_m\end{pmatrix}\end{aligned}$$

According to the definition of 1-norms, we know:

$$\|\mathbf{Ax}\|_1 = \sum_{i=1}^{m}|y_i| \quad \text{and} \quad \|\mathbf{x}\|_1 = \sum_{j=1}^{n}|x_j|$$

$$\begin{aligned}\frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1} = \frac{\sum_{i=1}^{m}|y_i|}{\sum_{j=1}^{n}|x_j|} &= \sum_{i=1}^{m}\frac{|a_{i1}x_1 + a_{i_2}x_2 + \cdots + a_{in}x_n|}{\sum_{j=1}^{n}|x_j|}\\ &\le \sum_{i=1}^{m}\frac{\sum_{j=1}^{n}|a_{ij}||x_j|}{\sum_{j=1}^{n}|x_j|}\\ &= \frac{\sum_{j=1}^{n}|x_j|\sum_{i=1}^{m}|a_{ij}|}{\sum_{j=1}^{n}|x_j|}\end{aligned}$$

Let $M = \max_{1\le j\le n}\sum_{i=1}^{m}|a_{ij}|$, we can derive:

$$\frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1} \le \frac{M\sum_{j=1}^{n}|x_j|}{\sum_{j=1}^{n}|x_j|} = M$$

It is obvious that the range of $\frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1}$ is a nonempty subset of real numbers, so its supremum exists and its supremum is M.

So $\|\mathbf{A}_1\| = \sup_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq 0} \frac{\|\mathbf{Ax}\|_1}{\|\mathbf{x}\|_1} = M = \max_{1\leq j\leq n}\sum_{i=1}^m |a_{ij}|$.

(b) According to the definition of ∞-norms, we know:

$$\|\mathbf{Ax}\|_\infty = \max_{1\leq i\leq m}\{|y_1|, |y_2|\cdots, |y_n|\}$$
$$\|\mathbf{x}\|_\infty = \max_{1\leq i\leq m}\{|x_1|, |x_2|\cdots, |x_n|\}$$

Let $\|\mathbf{x}\|_\infty = |x'|$, and thus $|x'| \geq |x_i|, i = 1, 2, \cdots, n$. According to triangle inequality, we have:

$$\begin{aligned}
|y_i| = |a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n| \leq \sum_{j=1}^n |a_{ij}||x_j| = |y_i'| \leq |x'|\sum_{j=1}^n |a_{ij}| \\
\frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty} &= \frac{\max_{1\leq i\leq m}\{|y_1|, |y_2|\cdots, |y_n|\}}{|x'|} \\
&\leq \frac{\max_{1\leq i\leq m}\{|x'|\sum_{j=1}^n |a_{ij}|, |x'|\sum_{j=1}^n |a_{ij}|\cdots, |x'|\sum_{j=1}^n |a_{ij}|\}}{|x'|} \\
&= \max_{1\leq i\leq m}\sum_{j=1}^n |a_{ij}| \\
\|\mathbf{A}\|_\infty &= \sup_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq 0}\frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty} \\
&= \max_{1\leq i\leq m}\sum_{j=1}^n |a_{ij}|
\end{aligned}$$

4. Exercise 4: Open and Closed Sets

4.1 Solution

(b) to (a): According to the definition of closed set: A set $F \subseteq R^n$ is closed if its complement set is open, it is obvious from (b) to (a).

(a) to (c): Assume $\exists\mathbf{x}' \notin C$, it satisfies$\forall\epsilon > 0, B_\epsilon(\mathbf{x}') \cap C \neq \emptyset$. We let:

$$(\mathbf{x_k}) = B_\epsilon(\mathbf{x}') \cap C$$
$$\exists N = 1, \text{s.t} \quad \text{when} k \geq N, \quad \|\mathbf{x_0} - \mathbf{x}'\| \leq \epsilon$$

According to the definition of limit point, we know $\mathbf{x}'$ is a limit point. Because C is a closed set that contains all of its limit points, we can derive that $\mathbf{x}' \in C$. This contradicts our premise.

(c) to (b) Let $\mathbf{x_0}$ be the limit point of C^C. It means:

$$\exists(\mathbf{x}_k) \subseteq C, \forall\epsilon > 0, \exists N, k > N, \|\mathbf{x_k} - \mathbf{x_0}\|_2 \leq \epsilon$$

In other words:

$$B_\epsilon(\mathbf{x_0}) \cap C^C \neq \emptyset$$

According to **(c)** we know:

$$\mathbf{x_0} \notin C^C$$

Thus we can say: C^C is an open set.

4.2 Solution

(a) $\exists\mathbf{x} = 1 \in [0, 1]$. $\forall\epsilon > 0, B_\epsilon(\mathbf{x}) \cap \mathbb{R} \not\subseteq [0, 1]$. Thus $[0, 1]$ is not an open set in $\mathbb{R}$.

$\forall 0 < \epsilon < 1 \quad \forall x \in [0, 1], B_\epsilon(x) \cap B \subset [0, 1]$, it is consistent with the definition of open set. Thus $[0, 1]$ is an open set in B.

On the other hand, $[0, 1] \backslash B = \{2\}$. $\forall x \in \{2\}, \forall\epsilon > 0, B_\epsilon(x) \cap B = 2 \in \{2\}$. So 2 is an open set in B. And according to the definition, $\{2\} \backslash B = [0, 1]$ is closed in B.

$[0, 1]$ is both open and closed in B.

(b) Let us first prove the sufficiency.

U is open in $\mathbb{R}^n$, it means:

$$\forall\mathbf{x} \in U, \exists\epsilon > 0, B_\epsilon(\mathbf{x}) \subset U$$

Because $C = A \cap U$, we have:

$$\forall\mathbf{x} \in C, \text{it satisfies} \quad \mathbf{x} \in U$$

According to the definition, we know:

$$\exists\epsilon > 0, \text{s.t} \quad B_\epsilon(\mathbf{x}) \subset U$$
$$B_\epsilon(\mathbf{x}) \cap A \subset U \cap A = C$$

C is open in A.

Next we prove the necessity.

If U is not open in $\mathbb{R}^n$, $U^C = \mathbb{R}^n \backslash U$ is open in $\mathbb{R}^n$:$\forall\mathbf{x_0} \in U^C \cap A, \exists\epsilon > 0, B_\epsilon(\mathbf{x_0}) \subset U^C$. Let ϵ_0 is the largest ϵ that meets this condition. If we let $d(\mathbf{x_1}, \mathbf{x_0}) = \epsilon_0$ and $\mathbf{x_1} \in A$, we can know $\mathbf{x_1} \notin U^C \cap A$. In other words, $\mathbf{x_1} \in U \cap A = C$.

Because C is open in A,

$$\exists\epsilon_1 > 0, B_{\epsilon_1}(\mathbf{x_1}) \subset U \cap A = C$$

In other words, $\forall \mathbf{x} \in B_{\epsilon_1}(\mathbf{x_1}), \mathbf{x} \in U$. But we easily find a $\mathbf{x_2} \in B_{\epsilon_1}(\mathbf{x_1}) \subset U$, it satisfies $d(\mathbf{x_2}, \mathbf{x_0}) < \epsilon_0$. That is:

$$\mathbf{x_2} \in B_{\epsilon_0}(\mathbf{x_0})$$
$$\mathbf{x_2} \in U^C$$

This contradicts $\mathbf{x_2} \in U$.

5. Exercise 5: Extreme Value Theorem

5.1 Solution

f is continuous, thus for $\epsilon = 1, \mathbf{x} \in c$,there exists a neighborhood of $\mathbf{x}$-$U_\mathbf{x}$, s.t $\forall \mathbf{x}' \in C \cap U_\mathbf{x}$, $|f(\mathbf{x}') - f(\mathbf{x})| < 1$. $|f(\mathbf{x}')| < |f(\mathbf{x})| + 1$. In other words, there exists upper bound of $|f|$ in $C \cap U_\mathbf{x}$.

The set of open sets $\{U_\mathbf{x} | \mathbf{x} \in C\}$ forms an open cover, which has finite subcovers: $\{U_1, U_2, \cdots, U_n\}$. Let's set the upper bound of $|f|$ in $C \cap U_n$ is $|f(\mathbf{x_n})| + 1$. There exists upper bound of $|f|$ in C: $M = 1 + \max\{|f(\mathbf{x}_1)|, |f(\mathbf{x}_2)|, \cdots, |f(\mathbf{x}_n)|\}$.

Let's set y_0 to be a limit point of $f(C)$. There exists a sequence $(y_n) \subseteq f(C)$, and $y_n \neq y_0$, $y_n \to y_0$. Let's set $f(\mathbf{x}_n) = y_n$. Because C is a compact set in $\mathbb{R}^n$, there exists a point $\mathbf{x_0}$ that a subsequence of $(\mathbf{x}_n)$ converges to: $\mathbf{x}_{n_k} \to \mathbf{x_0}$.

Because f is continuous, we have:

$$y_0 = \lim_{k \to \infty} y_{n_k} = \lim_{k \to \infty} f(\mathbf{x}_{n_k}) = f(\lim_{k \to \infty} \mathbf{x}_{n_k}) = f(\mathbf{x}_0) \in f(C)$$

$f(C)$ is closed and bounded, thus $f(C)$ is compact. According to the compactness, f can attain its maximum and minimum:

$$\exists \mathbf{a} \in C, f(\mathbf{a}) = \max f$$
$$\exists \mathbf{b} \in C, f(\mathbf{b}) = \min f$$
$$\forall \mathbf{x} \in C, f(\mathbf{a}) \leq f(\mathbf{x}) \leq f(\mathbf{b})$$

5.2 Solution

Useing the result of (5.1), we know $f([a, b])$ is bounded and closed. we can set the range of f is $[c, d]$. Besides, we know there exists some $x \in [a, b]$ that make f attain c and d.

6. Exercise 6: Basis and Coordinates

6.1 Solution

$\{\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_n}\}$ is a basis, thus $\forall \mathbf{x} \in V$:

$$\mathbf{x} = \beta_1 \mathbf{a_1} + \beta_2 \mathbf{a_2} + \cdots + \beta_n \mathbf{a_n}$$

The coefficients (β_k) are uniquely determined by $\mathbf{x}$. We let $\beta_k = \gamma_k\lambda_k$,

$$\mathbf{x} = \gamma_1(\lambda_1\mathbf{a_1}) + \gamma_2(\lambda_2\mathbf{a_2}) + \cdots + \gamma_n(\lambda_n\mathbf{a_n})$$

6.2 Solution

$\mathbf{P}$ is invertible, let's set $\mathbf{P}^{-1} = (p_{ij})$.$\mathbf{A} = \mathbf{BP^{-1}}$, that is:

$$\begin{aligned}
\mathbf{a_i} &= \sum_{k=1}^{n} p_{ki}\mathbf{b_k} \\
\forall \mathbf{x} \in V, \mathbf{x} &= \beta_1\mathbf{a_1} + \cdots + \beta_n\mathbf{a_n} \\
&= \beta_1\sum_{k=1}^{n} p_{k1}\mathbf{b_k} + \cdots + \beta_n\sum_{k=1}^{n} p_{kn}\mathbf{b_k} \\
&= (\beta_1 p_{11} + \cdots + \beta_n p_{1n})\mathbf{b_1} + \cdots + (\beta_1 p_{n1} + \cdots + \beta_n p_{nn})\mathbf{b_n} \\
&= \alpha_1\mathbf{b_1} + \alpha_2\mathbf{b_2} + \cdots + \alpha_n\mathbf{b_n}
\end{aligned}$$

Therefore we know that $\{\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_n}\}$ is the basis of V.

6.3 Solution

(a) Using the result of (6.1), we know $\gamma_k = \frac{x_k}{\lambda_k}$. It means the coordinates of $\mathbf{v}$ is $(\frac{x_1}{\lambda_1}, \frac{x_2}{\lambda_2}, \cdots, \frac{x_n}{\lambda_n})$.

(b) Under $\{\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_n}\}$, the coordinates is $(1, 1, \cdots, 1)$. Under $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, \cdots, \lambda_n\mathbf{a_n}\}$, the coordinates is $(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_n})$.

6.4 Solution

(a) $\mathbf{v} = y\mathbf{b} - x\mathbf{c}$. Therefore the coordinates under the $\mathbf{c}, \mathbf{b}$ is $(-x, y)$. It is not unique. There may be other coordinates under other basis.

(b)

$$\begin{aligned}
(x, y) &= (x' - z', y') \\
y' &= y, x = x' - z'
\end{aligned}$$

It means: $\mathbf{v} = x'\mathbf{a} + y\mathbf{b} + (x' - x)\mathbf{c}, x' \in \mathbb{R}$.

(c)

$$\begin{aligned}
\|(x', y', z')\|_1 &= |x'| + |y'| + |z'| \\
&= |x'| + |x' - x| + |y| \\
&\geq |x' - (x' - x)| + |y| \\
&= |x| + |y|
\end{aligned}$$

When $x' = 0$ or $x' = x$, it attains the minimum.

7. Exercise 7: Derivatives with Matrices

7.1 Solution

(a) Let $L = \mathbf{a}^T$,

$$\forall \mathbf{x_0} \in \mathbb{R}^n, \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \frac{\|\mathbf{a^T x} - \mathbf{a^T x_0} - \mathbf{a^T}(\mathbf{x} - \mathbf{x0})\|_2}{\|\mathbf{x} - \mathbf{x_0}\|\|2} = 0$$

Thus f is differentiable and $f'(\mathbf{x}) = \mathbf{a}^T$.

(b) Let $L = 2\mathbf{x}^\top$,

$$\begin{aligned}
\forall \mathbf{x_0} \in \mathbb{R}^n, \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \frac{\|\mathbf{x}^\top \mathbf{x} - \mathbf{x_0}^\top \mathbf{x_0} - \mathbf{2x}^\top (\mathbf{x} - \mathbf{x_0})\|_2}{\|\mathbf{x} - \mathbf{x_0}\|_2} &= \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \frac{\|(\mathbf{x} - \mathbf{x_0})^\top (\mathbf{x} - \mathbf{x_0})\|_2}{\|\mathbf{x} - \mathbf{x_0}\|_2} \\
&= \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \frac{\|\mathbf{x} - \mathbf{x_0}\|_2 \|\mathbf{x} - \mathbf{x_0}\|_2}{\|\mathbf{x} - \mathbf{x_0}\|_2} \\
&= \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \|\mathbf{x} - \mathbf{x_0}\|_2 \\
&= 0
\end{aligned}$$

Thus f is differentiable and $f'(\mathbf{x}) = 2\mathbf{x}^T$.

(c) Let $L = -2(\mathbf{y} - \mathbf{Ax})^\top \mathbf{A}$,

$$\begin{aligned}
f(\mathbf{x}) - f(\mathbf{x_0}) - L(\mathbf{x} - \mathbf{x_0}) &= (\mathbf{y} - \mathbf{Ax})^\top (\mathbf{y} - \mathbf{Ax}) - (\mathbf{y} - \mathbf{Ax_0})^\top (\mathbf{y} - \mathbf{Ax_0}) \\
&\quad + 2(\mathbf{y} - \mathbf{Ax})^\top \mathbf{A}(\mathbf{x} - \mathbf{x_0}) \\
&= \mathbf{x}^\top \mathbf{A}^\top \mathbf{Ax} - \mathbf{x_0}^\top \mathbf{A}^\top \mathbf{Ax_0} - 2\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}(\mathbf{x} - \mathbf{x_0}) \\
&= -\mathbf{x}^\top \mathbf{A}^\top \mathbf{A}(\mathbf{x} - \mathbf{x_0}) + (\mathbf{x} - \mathbf{x_0})^\top \mathbf{A}^\top \mathbf{Ax_0} \\
&= -(\mathbf{x} - \mathbf{x_0})^\top \mathbf{A}^\top \mathbf{A}(\mathbf{x} - \mathbf{x_0}) \\
&= -\|\mathbf{A}(\mathbf{x} - \mathbf{x_0})\|_\mathbf{2}^\mathbf{2}
\end{aligned}$$

$$\begin{aligned}
\forall \mathbf{x_0} \in \mathbb{R}^n, \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \frac{\|f(\mathbf{x}) - f(\mathbf{x_0}) - L(\mathbf{x} - \mathbf{x_0})\|_2}{\|\mathbf{x} - \mathbf{x_0}\|_2} &= \lim_{\mathbf{x} \to \mathbf{x_0}; \mathbf{x} \neq \mathbf{x_0}} \frac{\|\mathbf{A}(\mathbf{x} - \mathbf{x_0})\|_\mathbf{2}^\mathbf{2}}{\|\mathbf{x} - \mathbf{x_0}\|_2} \\
&= 0
\end{aligned}$$

Thus f is differentiable. $f'(\mathbf{x}) = -2(\mathbf{y} - \mathbf{Ax})^\top \mathbf{A}$.

7.2 Solution

Let $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ be a function, $(\mathbf{X_0}) \in \mathbb{R}^{n \times n}$ is a point, and let $L : \mathbb{R}^{n \times n} \to \mathbb{R}$ be a linear transformation. We say that f is $differentiable$ at $(\mathbf{X_0})$ with derivative L if we have

$$\lim_{\mathbf{X} \to \mathbf{X_0}; \mathbf{X} \neq \mathbf{X_0}} \frac{\|f(\mathbf{X}) - f(\mathbf{X_0}) - L(\mathbf{X} - \mathbf{X_0})\|_2}{\|\mathbf{X} - \mathbf{X_0}\|_2}$$

We denote this derivative L by $f'(\mathbf{X_0})$.

7.3 Solution

$f(\mathbf{x}') = \mathbf{A}^\top$

8. Exercise 8: Rank of Matrices

8.1 Solution

(a) Suppose $\mathbf{PAQ} = \operatorname{diag}(\mathbf{I_r}, 0)$. It is obvious that $\mathbf{P}^\top\mathbf{A}^\top\mathbf{Q}^\top = \operatorname{diag}(\mathbf{I_r}, 0)$. Thus $\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top) = r$.

(b) We let $\mathbf{C} = \mathbf{AB}$,$\mathbf{A} = (a_{ij})_{m\times n}$ and the row vectors of $\mathbf{B}$ be: $\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_n}$ and the row vectors of $\mathbf{C}$ be: $\mathbf{c_1}, \mathbf{c_2}, \cdots, \mathbf{c_m}$.

$$\mathbf{c_i} = a_{i1}\mathbf{b_1} + a_{i2}\mathbf{b_2} + \cdots + a_{in}\mathbf{b_n}, i = 1, 2, \cdots, m$$

It means that the vector groups $\{\mathbf{c_1}, \mathbf{c_2}, \cdots, \mathbf{c_m}\}$ can be represented by the combinations of $\{\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_n}\}$. Thus $\mathbf{rand}(\mathbf{AB}) \leq \mathbf{rank}(\mathbf{B})$. On the other hand $\mathbf{rank}(\mathbf{AB}) = \mathbf{rank}(\mathbf{B}^\top\mathbf{A}^\top) \leq \mathbf{rank}(\mathbf{A}^\top) = \mathbf{rank}(\mathbf{A})$.

When $\mathbf{B}$ is Identity matrix, the equal sign is true.

8.2 Solution

(a) Let's set the column vectors of $\mathbf{A}$ are $\{\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_n}\}$. According to the definition of rank, $\mathbf{rank}(\mathbf{A})$ equals the number of vectors in the maximal independent group of vector groups $\{\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_n}\}$, denoted by r.

Suppose the maximal independent group is $\{\mathbf{b_1}, \mathbf{b_2}, \cdots, \mathbf{b_r}\}$, it is a basis of column space. And we know the dimension of a linear space equals the number of basis. $\dim(\mathcal{C}(\mathbf{A})) = r = \mathbf{rank}(\mathbf{A})$.

(b) Suppose $\mathbf{x}, \mathbf{y} \in \mathcal{N}(\mathbf{A})$, $\mathbf{Ax} = \mathbf{0}, \mathbf{Ay} = \mathbf{0}$. It means $\mathbf{A}(\mu\mathbf{x} + \lambda\mathbf{y}) = \mathbf{0}$. Thus $\mu x + \lambda y \in \mathcal{N}(\mathbf{A})$. $\mathcal{N}(\mathbf{A})$ is a subspace of $\mathbb{R}^n$.

We do elementary row transformation to system of linear equations $\mathbf{Ax} = \mathbf{0}$ to get $\mathbf{Jx} = \mathbf{0}$. $\mathbf{J}$ is a matrix in ladder form. And we know $\mathbf{rank}(\mathbf{J}) = \mathbf{rank}(\mathbf{A}) = r$. The solution of $\mathbf{Ax} = \mathbf{0}$ is:

$$\mathbf{x} = t_1\alpha_\mathbf{1} + \cdots + t_{n-r}\alpha_\mathbf{n-r}$$

It is not difficult to find $\{\alpha_\mathbf{1}, \alpha_\mathbf{2}, \cdots, \alpha_\mathbf{n-r}\}$ is linear independent. Thus $\dim()\mathcal{N}(\mathbf{A}) = n - r = n - \mathbf{rank}(\mathbf{A})$.

9. Exercise 9: Linear Equations

9.1 Solution

(a)

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

There exists unique solution:$\mathbf{x} = (1, 0)^\top$.

(b)

$$\mathbf{W} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

There aren't any solutions.

(c)

$$\mathbf{W} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{y} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

Suppose $\mathbf{x} = (x_1, x_2, x_3)^\top$. We can get $x_1 + x_3 = 1$ and $x_2 = 1$. It means there exists more than one solution.

9.2 Solution

Suppose the columns of $\mathbf{X}$ are $\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_d}$, $\mathbf{w} = (w_1, w_2, \cdots, w_d)^\top$. $\mathbf{Xw} = \mathbf{y}$ can be written as:

$$w_1\mathbf{a_1} + w_2\mathbf{a_2} + \cdots + w_d\mathbf{a_d} = \mathbf{y}$$

The system of linear equations has solution $\Leftrightarrow \mathbf{y} \in < \mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_d} > \Leftrightarrow \mathbf{rank}(\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_d}) = \mathbf{rank}(\mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_d}, \mathbf{y}) \Leftrightarrow \mathbf{rank}(\mathbf{X}) = \mathbf{rank}(\mathbf{X}, \mathbf{y})$.

The solution is unique $\Leftrightarrow \mathbf{a_1}, \mathbf{a_2}, \cdots, \mathbf{a_d}$ are linear independent and $\mathbf{rank}(\mathbf{X}) = \mathbf{rank}(\mathbf{X}, \mathbf{y})$ $\Leftrightarrow \mathbf{rank}(\mathbf{X}) = \mathbf{rank}(\mathbf{X}, \mathbf{y}) = d$.

9.3 Solution

10. Exercise 10: Properties of Eigenvalues and Singular Values

10.1 Solution

Suppose $\mathbf{A} = \mathbf{P^{-1}\Lambda P} \;=\; \mathbf{A} = \mathbf{P}^\top\mathbf{\Lambda P}$. $\mathbf{P}$ is orthogonal matrix and $\mathbf{\Lambda}$ is eigenvalue matrix.

$$\begin{aligned}\mathbf{x}^\top\mathbf{Ax} &= (\mathbf{Px})^\top\mathbf{\Lambda}(\mathbf{Px}) \\ &\leq \lambda_{max}(\mathbf{Px})^\top(\mathbf{Px}) \\ &= \lambda_{max}\|\mathbf{x}\|_\mathbf{2}^\mathbf{2}\end{aligned}$$

Therefore, $\lambda_{max}(\mathbf{A}) = \sup_{\mathbf{x}\in\mathbb{R}^n;\mathbf{x}\neq\mathbf{x_0}} \frac{\mathbf{x}^\top\mathbf{Ax}}{\mathbf{x}^\top\mathbf{x}}$. On the other hand,

$$\begin{aligned}\mathbf{x}^\top\mathbf{Ax} &= (\mathbf{Px})^\top\mathbf{\Lambda}(\mathbf{Px}) \\ &\geq \lambda_{min}(\mathbf{Px})^\top(\mathbf{Px}) \\ &= \lambda_{min}\|\mathbf{x}\|_\mathbf{2}^\mathbf{2}\end{aligned}$$

Therefore, $\lambda_{min}(\mathbf{A}) = \inf_{\mathbf{x}\in\mathbb{R}^n;\mathbf{x}\neq\mathbf{x_0}} \frac{\mathbf{x}^\top\mathbf{Ax}}{\mathbf{x}^\top\mathbf{x}}$.