Machine Learning HW1

Liu Pei PB22061259

September 2024

1 Exercise 1: Bolzano-Weierstrass Theorem

1.1

Prove: Suppose that u is not a upper bound of C,but we know that there is an upper bound of C,let's just call it u',so we have

$$u' < u, u' = supC$$

let $\varepsilon = \frac{u-u'}{2}$, so

$$u - \varepsilon = \frac{u+u'}{2} > u'$$

because $a \in C$, so $a \leq u'$, so it is impossible that

$$a > u - \varepsilon = \frac{u+u'}{2} > u'$$

so the hypothesis is wrong,which means u = supC

1.2

Prove:

1. Divide the interval $[a, b]$ into two subintervals $[a, \frac{a+b}{2}]$ and $[\frac{a+b}{2}, b]$. By the Pigeonhole Principle, at least one of these subintervals contains infinitely many terms of the sequence $\{a_n\}$. Denote this subinterval by I_1.

2. Repeat the process of bisecting I_1 into two smaller subintervals and selecting the one that contains infinitely many terms of $\{a_n\}$. Denote this new subinterval by I_2.

3. Continue this process indefinitely, resulting in a nested sequence of closed intervals $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$, where each I_k has length $|I_k| = \frac{1}{2^{k-1}}|I_1|$ and contains infinitely many terms of $\{a_n\}$.

4. According to the Nested Interval Theorem, there exists a unique real number c such that $c \in I_k$ for all k.

5. For each k, choose a term a_{n_k} from the sequence $\{a_n\}$ such that $a_{n_k} \in I_k$. Since $|I_k| \to 0$ as $k \to \infty$, by the Squeeze Theorem, it follows that $a_{n_k} \to c$.

6. Therefore, the subsequence $\{a_{n_k}\}$ of $\{a_n\}$ is convergent with limit c.

2 Exercise 2: Limit and Limit Points

2.1

1. **Boundedness:** Suppose $\{x_n\}$ converges to x. By definition, for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $n \geq N$, $\|x_n - x\| < \epsilon$. Choose $\epsilon = 1 + \|x\|$. Then there exists an N such that for all $n \geq N$,

$$\|x_n\| \leq \|x_n - x\| + \|x\| < 1 + \|x\| + \|x\| = 1 + 2\|x\|.$$

For $n < N$, the terms $x_1, x_2, \ldots, x_{N-1}$ are all finite vectors in $\mathbb{R}^n$, so they are automatically bounded by some constant M (e.g., take $M = \max\{\|x_1\|, \|x_2\|, \ldots, \|x_{N-1}\|, 1+2\|x\|\}$). Therefore, for all $n \in \mathbb{N}$, $\|x_n\| \leq M$, showing that $\{x_n\}$ is bounded.

2. **Unique Limit Point:** Suppose $\{x_n\}$ converges to both x and y. Then for every $\epsilon > 0$, there exist $N_1, N_2 \in \mathbb{N}$ such that for all $n \geq N_1$, $\|x_n - x\| < \epsilon$ and for all $n \geq N_2$, $\|x_n - y\| < \epsilon$. Taking $N = \max\{N_1, N_2\}$, we have for all $n \geq N$,

$$\|x - y\| \leq \|x - x_n\| + \|x_n - y\| < 2\epsilon.$$

Since ϵ is arbitrary, we can make it as small as we want, implying $\|x-y\| = 0$, and thus $x = y$. Therefore, the limit is unique.

Second Direction: If $\{x_n\}$ is bounded and has a unique limit point x, then $\{x_n\}$ converges to x.

2. **Convergence:** Suppose $\{x_n\}$ is bounded and has a unique limit point x. Since $\{x_n\}$ is bounded, it is contained in some closed ball $B(0, R)$ centered at the origin with radius R. By the Bolzano-Weierstrass Theorem (or a generalization to $\mathbb{R}^n$), every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence. Let $\{x_{n_k}\}$ be a convergent subsequence of $\{x_n\}$ that converges to y. Since x is the unique limit point of $\{x_n\}$, it must be that $y = x$.

Now, we need to show that the entire sequence $\{x_n\}$ converges to x. Suppose not. Then there exists an $\epsilon > 0$ and a subsequence $\{x_{m_j}\}$ such that $\|x_{m_j} - x\| \geq \epsilon$ for all j. However, since $\{x_{m_j}\}$ is also a bounded sequence in $\mathbb{R}^n$, it must have a convergent subsequence, say $\{x_{m_{j_k}}\}$, which converges to some z. Since x is the unique limit point of $\{x_n\}$, it must be that $z = x$. But this contradicts the fact that $\|x_{m_{j_k}} - x\| \geq \epsilon$ for all k. Therefore, our assumption that $\{x_n\}$ does not converge to x is false, and hence $\{x_n\}$ converges to x.

2.2

(a)

1. assume that x = 2 is not an isolated point of C , which means that there is a sequence $\{x_n\}$ in C such that $\{x_n\} \to 2$,so

$$\forall \epsilon > 0, \exists N, n > N, \|x_n - 2\| < \epsilon$$

   let $\epsilon = 0.5$,so we have

$$\forall n > N, x_n \in [1.5, 2.5]$$

   which against

$$\{x_n\} \in C, x_n \neq 2$$

   so the hypothesis is wrong , x=2 is an isolated point of C

2. $C' = [0, 1]$**:**
   Let a $\in C'$,so that $\forall \varepsilon > 0$,we have

$$a - \varepsilon \in [-\varepsilon, 1 - \varepsilon]$$

   and

$$a + \varepsilon \in [\varepsilon, 1 + \varepsilon]$$

   so the neighborhood of a is $[a - \varepsilon, a + \varepsilon]$
   which means every neighborhood of a contains a point $a' \in C$
   so C' = [0,1]

(b)
let's assume that C' is open
it is easy to know that

$$\exists b, \forall \epsilon > 0, b \in \mathbb{R}^n \backslash C', b - \epsilon \in C' (or : b + \epsilon \in C' \text{we just choose one to discuss})$$

because $b - \epsilon \in C'$, so there is a sequence $\{x_n\}$ in C such that $\{x_n\} \to b - \epsilon$
such that $\forall \epsilon_1 > 0, \exists N > 0,$ when $n > N$
we have

$$\|x_n - b + \epsilon\| < \epsilon_1$$

such that

$$\|\|x_n - b\| - \epsilon\| \leq \|x_n - b + \epsilon\| < \epsilon_1$$

so we have

$$|\|x_n - b\| - \epsilon \leq \|\|x_n - b\| - \epsilon\| \leq \|x_n - b + \epsilon\| < \epsilon_1$$

so

$$\|x_n - b\| < \epsilon + \epsilon_1$$

which means that there is a sequence $\{x_n\}$ in C such that $\{x_n\} \to b$
so b $\in C'$,which against our assumption
so C' is closed

3 Exercise 3: Norms

3.1

a

1. f is nonnegative: because $|x_i| \geq 0$, so $|x_i|^p \geq 0$,so $\| \mathbf{x} \|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p} \geq 0$

2. f is definite: if

$$\| \mathbf{x} \|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} = 0$$

   such that only if $\sum_{i=1}^{n} |x_i|^p = 0$
   for $|x_i|^p \geq 0$ so that only if $|x_i| = 0$,we can get $\| \mathbf{x} \|_p = 0$

3. f is homogeneous - in this section ,we are going to prove :

$$f(t\mathbf{x}) = |t| f(\mathbf{x}), \text{for all } \mathbf{x} \in \mathbb{R}^n \text{ and } t \in \mathbb{R};$$

$$\| \mathbf{tx} \|_p = \left(\sum_{i=1}^{n} |tx_i|^p \right)^{1/p} = \left(|t|^p \sum_{i=1}^{n} |tx_i|^p \right)^{1/p} = |t| \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

   and that‘s it

4. f satisfies the triangle inequality-in this section ,we are going to prove :

$$f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y}), \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

   which means this :

$$\left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p \right)^{\frac{1}{p}}$$

   as we all know ,this is the famous Minkowski Inequality ,of cause right
   To sum up,the l_p norm is a norm.

b

we are going to prove:

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|.$$

here, we can try to use Squeeze Theorem:
the right side:

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \lim_{p\to\infty} \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \le \lim_{p\to\infty} \left(\sum_{i=1}^{n} \max_{1\le i\le n} |x_i|^p\right)^{1/p} = \lim_{p\to\infty} \left(n \max_{1\le i\le n} |x_i|^p\right)^{1/p} =$$

$$\lim_{p\to\infty} n^{1/p} \max_{1\le i\le n} |x_i| = \max_{1\le i\le n} |x_i|$$

and the left side: considering $|x_i| \ge 0$

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \lim_{p\to\infty} \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \ge \lim_{p\to\infty} \left(\max_{1\le i\le n} |x_i|^p\right)^{1/p} = \max_{1\le i\le n} |x_i|$$

so ,by Squeeze Theorem ,we can say

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty = \max_{1\le i\le n} |x_i|.$$

3.2 Operator norms

a

Let $x = (\lambda_1, \lambda_2, ..., \lambda_n)^T$,so ,$\|\mathbf{x}\|_1 = \sum_{j=1}^{n} |\lambda_j|$
Let A_i be the column vector, so the A can be writed as $(A_1, A_2, ..., A_n)$
then

$$\|Ax\|_1 = \|\sum_{j=1}^{n} A_j\lambda_j\|_1 = \sum_{j=1}^{n} \lambda_j \|A_j\|_1 = \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{m} |a_{ij}|$$

so,we can know that

$$\|\mathbf{A}\|_1 = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\ne\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_1}{\|\mathbf{x}\|_1} = sup \frac{\sum_{j=1}^{n} \lambda_i \sum_{i=1}^{m} |a_{ij}|}{\sum_{j=1}^{n} |\lambda_i|}$$

thinking that

$$\frac{\sum_{j=1}^{n} \lambda_j \sum_{i=1}^{m} |a_{ij}|}{\sum_{j=1}^{n} |\lambda_i|} \le \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$$

so,we have

$$\|\mathbf{A}\|_1 = \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|.$$

b

Let $x = (\lambda_1, \lambda_2, ..., \lambda_n)^T$,so ,$\|\mathbf{x}\|_\infty = \max_{1\le j\le n} |\lambda_j|$

Let A_i be the column vector, so the A can be writed as $(A_1, A_2, ..., A_n)$
then

$$\|Ax\|_\infty = \|\sum_{j=1}^{n} A_j\lambda_j\|_\infty = max_{1\le i\le m}\sum_{j=1}^{n}\lambda_j a_{ij}$$

so,we can know that

$$\|\mathbf{A}\|_\infty = \sup_{\mathbf{x}\in\mathbb{R}^n,\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{Ax}\|_\infty}{\|\mathbf{x}\|_\infty} = sup\frac{max_{1\le i\le m}\sum_{j=1}^{n}\lambda_j a_{ij}}{\max_{1\le j\le n}|\lambda_j|}$$

thinking that

$$\frac{max_{1\le i\le m}\sum_{j=1}^{n}\lambda_j a_{ij}}{\max_{1\le j\le n}|\lambda_j|} \le \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|.$$

so

$$\|\mathbf{A}\|_\infty = \max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}|.$$

4 Exercise 4: Open and Closed Sets

4.1

4.1.1 $a \to b$

if cl C = C,assume the complement of C is not open:

the complement of C is not open,there is a point x in the complement of C,$\forall r$,whose $B_r(\mathbf{x})$ must contains some points y $\in C$
so we can say

$$\forall r > 0, B_r(\mathbf{x}) \text{ contains some y}, y \in C$$

what if cl C = C ,which means

$$\forall y \text{ in C }, x \text{ is a limit point of C}$$

according to

Definition 17. A vector x $\in \mathbb{R}^n$ is a limit (cluster/accumulation) point of a set $S \subseteq \mathbb{R}^n$ if

every neighborhood of x contains a point $x' \neq$ x such that $x' \in S$.

here we get x ,whose every r-neighborhood contains a point y$\neq$x,such that y$\in C$
so , y in the complement of C is a limit point of C,which against cl C = C.
To sum up ,the complement of C is open

4.1.2 $b \to a$

if the complement of C (let‘s call it C')is open,we can get there exists an $r_1 > 0$ such that

$$\forall x \in C', B_r(\mathbf{x}) \subseteq C'$$

the $B_r(\mathbf{x})$ upside don‘t contain any point of C, so that $\forall x \in C'$, x is not a limit point of C

we choose a point y in C,pay attention to it‘s r-neighborhood:
if there exist $r_2 > 0$,$B_r(\mathbf{y})$ don’t contain a point $y' \neq$ y in C
we can say there exist r > 0,such that

$$B_r(\mathbf{y}) \subseteq C', \text{except y}$$

let $r = min(r_1, r_2)$, we have :

$$\forall x \in C', B_r(\mathbf{x}) \subseteq C'$$

$$\exists y \in C, B_r(\mathbf{y}) \subseteq C', \text{except y}$$

here we can easily find the absurd:

$$\forall y' \in B_r(\mathbf{y}), y' \in C', B_r(\mathbf{y'}) \text{ contains y}$$

which against

$$\forall x \in C', B_r(\mathbf{x}) \subseteq C'$$

so,there do not exist a point y in C ,whose r-neighborhood don‘t contain a point $y' \neq$ y in C,which means every point in C, it‘s every r-neighborhood contain a point $y' \neq$ y in C

so every point of C is a limit point of set C,which means cl C = C ,and that is the set C is closed.

4.1.3 $b \to c$

The complement of C(let‘s call it C') is open:we get there exist an $\epsilon > 0, \forall x \in C', B_\epsilon(\mathbf{x}) \subseteq C'$

so we can know $\forall y \in C'$,$B_\epsilon(\mathbf{y}) \cap C = \emptyset$ for some $\epsilon > 0$
so there we can‘t find any y in C‘ that $B_\epsilon(\mathbf{y}) \cap C \neq \emptyset$ for the ϵ we say above

so we have if $B_\epsilon(\mathbf{x}) \cap C \neq \emptyset$ for every $\epsilon > 0$, then $\mathbf{x} \notin C'$, then $\mathbf{x} \in C$.

4.1.4 $c \to a$

If $B_\epsilon(\mathbf{x}) \cap C \neq \emptyset$ for every $\epsilon > 0$, then $\mathbf{x} \in C$.
If $B_\epsilon(\mathbf{x}) \cap C \neq \emptyset$ for every $\epsilon > 0$:

1. if $B_\epsilon(\mathbf{x}) \cap C \neq x$,x is the limit point of C
2. if $B_\epsilon(\mathbf{x}) \cap C = x$,x is in C

we can easily exclude the second case:
if $B_\epsilon(\mathbf{x}) \cap C = x$,we can find a y(different with x) in$B_\epsilon(\mathbf{x})$,such that $B_\epsilon(\mathbf{y}) \cap C \neq \emptyset$,however,y is not in C,which against the precondition;

so we can say If $B_\epsilon(\mathbf{x}) \cap C \neq \emptyset$ for every $\epsilon > 0$,,then x in C,we have x is the limit point of C,and x in C,that is cl C = C.

4.2

4.2.1 a

first ,we know 0 is in [0,1] , but for $\forall\epsilon > 0, B_\epsilon(\mathbf{0}) = (-\epsilon, \epsilon), B_\epsilon(\mathbf{0}) \cap \mathbb{R} = (-\epsilon, \epsilon) \not\subset [0, 1]$,so [0, 1] is not an open set in $\mathbb{R}$.

$\exists\epsilon = 0.5, \forall x \in [0, 1], B_\epsilon(\mathbf{x}) = (-0.5 + x, 0.5 + x), B_\epsilon(\mathbf{x}) \cap B \subset [0, 1]$,so [0, 1] is open in B;

B \[0,1] = {2},so the neighborhood of 2 in B is some points in [0,1],so $C = \{2\}, B_\epsilon(\mathbf{x}) \cap B = \emptyset \subset C$ for some$\epsilon > 0$,so B \[0,1] is open, which means [0,1] is closed in B

To sum up,it is both open and closed in B.

4.2.2 b

1. If $C = A \cap U$ for some open set U in R^n, then C is open in A.
   we know that U is open in R , so $\forall x \in (U \cap A), x \in U, \exists\epsilon > 0, B_\epsilon(\mathbf{x}) \subset U$
   Considering $B_\epsilon(\mathbf{x}) \cap A$,because $B_\epsilon(\mathbf{x}) \subset U$,so,$(B_\epsilon(\mathbf{x}) \cap A) \subset (U \cap A) = C$
   so we get set C is open in A,if $C = A \cap U$ for some open set U in R^n

2. If C in open in A,such that $C = A \cap U$, where U is open in $\mathbb{R}^n$.
   Because C is open in A, so

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

   let $U = \bigcup_{x \in C} B_{\epsilon_x}(x)$,Since each $B_{\epsilon_x}(x)$ is open in $\mathbb{R}^n$, U is also open in $\mathbb{R}^n$ (union of open sets is open).

   (a) Define $U = \bigcup_{x \in C} B_{\epsilon_x}(x)$. Since each $B_{\epsilon_x}(x)$ is open in $\mathbb{R}^n$, U is also open in $\mathbb{R}^n$ (union of open sets is open).

(b) To show $C = A \cap U$:

i. $C \subset A \cap U$: For any $x \in C$, there exists ϵ_x such that $x \in B_{\epsilon_x}(x) \subset U$. Since $x \in C \subset A$, $x \in A \cap U$.

ii. $A \cap U \subset C$: For any $y \in A \cap U$, there exists $x \in C$ and ϵ_x such that $y \in B_{\epsilon_x}(x)$. Since $B_{\epsilon_x}(x) \cap A \subset C$, $y \in C$.

To sum up, If C in open in A,such that $C = A \cap U$, where U is open in $\mathbb{R}^n$.

5 Exercise 5: Extreme Value Theorem and Fixed Point

5.1

If f(0) = 0 or f(1) = 1, it is obvious that there is an x such that f(x) = x;
If not,let g(x) = f(x) - x,because f is a continuous ,so g is continuous too.
for $g(0) = f(0) > 0, g(1) = f(1) - 1 < 0$,and g is continuous,so

$$\exists x \in [0, 1], \text{ such that } g(x) = f(x) - x = 0.$$

that is

$$\forall f : [0, 1] \to [0, 1], \exists x \in [0, 1], \text{such that} f(x) = x.$$

5.2

f(x) = x^2 is a good example:

$$f(x) = x^2, \text{ when } x \in (0, 1), f \in (0, 1), \text{ and for all x} \in (0, 1), f(x) \neq x$$

5.3

First we prove that f is an one to one map.$\forall x, y \in [0, 1]$,if $f(x) = f(y)$,then

$$x = f^n(x) = f^{n-1}(f(x)) = f^{n-1}(f(y)) = f^n(y) = y,$$

Hence f is an injective map.Next we show that f is monotonic.if

$$\exists x, y \in [0, 1], x < y, f(x) \geqslant f(y),$$

because the range of f is $[0, 1]$, so $f(0) = 0 \leqslant f(y), f(x) \leqslant 1 = f(1)$
$\Longrightarrow x \neq 0, y \neq 1$.because f is an injection,$x \neq y \Longrightarrow f(x) > f(y), f(x) \neq 0 \Longrightarrow 0 = f(0) < f(x), f(x) > f(y)$.According to Rolle's Theorem,there exists $z_1 \in (0, x), z_2 \in (x, y)$,s.t.$f(z_1) = f(z_2) = \frac{f(x)+f(y)}{2}$,which contradicts with the injectivity of f.Thus f is monotonic.If $\exists x \in [0, 1], f(x) \neq x$,without lose of generality,we can assume that $f(x) < x$.Then

$$f^n(x) = f^{n-1}(f(x)) < f^{n-1}(x) < ... < f(x) < x,$$

which contradicts with the condition.So $\forall x \in [0, 1], f(x) = x$.

5.4

Let $g = \lambda f$.Note that (a) is a special case of (b).Because f is continuous,f is bounded,i.e $\exists M > 0$,s.t.$|f(x)| \leqslant M$. Let $\lambda = \frac{M}{2}$to make sure that $g(x) \in (0, 1)$, thus $g^n(x)$ is well-defined for any $n \in \mathbb{N}$.if $g(x) \geqslant x$,because f is non-decreasing,g is also non-decreasing,$x \leqslant g(x) \leqslant g(g(x)) \leqslant ...g^n(x) \leqslant ... \Longrightarrow \{g^n(x)\}$is a non-decreasing sequence and it is bounded,so the sequence must converge to a point $x_0 \in [0, 1]$.So

$$g(x_0) = g(\lim_{n\to\infty} f^n(x)) = \lim_{n\to\infty} g^{n+1}(x) = x_0.$$

Hence the sequence satisfy the case(b).Similarly for the case that $g(x) < x$.

6 Exercise 6: Linear Space

6.1

We can easily prove that the vector addition and scalar multiplication in this space satisfy the eight axioms.

$$u = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

$$v = b_n x^n + b_{n-1}x^{n-1} + \cdots + b_1 x + a_0$$

and

$$\lambda(u+v) = \lambda u+\lambda v = \lambda(a_n x^n+b_n x^n)+\lambda(a_{n-1}x^{n-1}+b_{n-1}x^{n-1})\cdots+\lambda(a_1 x+b_1 x)+\lambda(a_0+b_0)$$

so $P_n[x]$ is a linear space.

6.2

a

Assume there are two zero vectors v_1, v_2, Then $v_1 \xlongequal{v_1\text{is zero}} v_1 + v_2 \xlongequal{v_2\text{is zero}} v_2, \Longrightarrow$ zero vector is uique.

b

Just like the part a ,we can assume there are two vectors v_1, v_2 they are both additive inverse of u. So we have $v_1 = v_1 + (u + v_2) = (v_1 + u) + v_2 = v_2$.Next,we will show that $0 \cdot v = \mathbf{0}$. Let $(0 \cdot v)'$ denote the additive inverse of $0 \cdot v$.Then $0 \cdot v + 0 \cdot v = (0+0) \cdot v = 0 \cdot v \Longrightarrow ((0 \cdot v)' + 0 \cdot v) + 0 \cdot v = (0 \cdot v)' + 0 \cdot v \Longrightarrow 0 \cdot v = \mathbf{0}$. $v' + v = (-1) \cdot v + v = (1 - 1) \cdot v = 0 \cdot v = \mathbf{0} \Longrightarrow v' = (-1) \cdot v$ is the unique inverse of v,we denote it as $-v$.

c

We only need to prove that$\lambda \cdot \mathbf{0} = \mathbf{0}$,for the rest part of the problem has been proved in the proof process of (b). Let v' denote the additive inverse of v. Then $\lambda \cdot \mathbf{0} = \lambda(v + v') = \lambda(v + (-1) \cdot v) = \lambda \cdot v + \lambda \cdot ((-1) \cdot v) = \lambda \cdot v + (\lambda \cdot (-1)) \cdot v = (\lambda - \lambda) \cdot v = 0 \cdot v = \mathbf{0}$.

d

If $\lambda \neq 0$,there exists a unique multiply inverse of λ in F,we denote it as λ^{-1}.Thus,$v = 1 \cdot v = (\lambda^{-1} \cdot \lambda)v = \lambda^{-1}(\lambda \cdot v) = \lambda^{-1} \cdot \mathbf{0} \xlongequal{(c)} \mathbf{0}$.

6.3

A linear space can one vector,finite vectors,countable infinite vectors or uncountable infinite vectors.A linear space at least have one subspace,which is itself.The linear space$\{\mathbf{0}\}$ has one subspace.

7 Exercise 7: Basis and Coordinates

7.1

if scalars $\mu_1, ..., \mu_n$ satisfy that $\sum_{i=1}^{n} \mu_i(\lambda_i\mathbf{a_i}) = 0 \Longrightarrow \sum_{i=1}^{n}(\mu_i\lambda_i)\mathbf{a_i} = 0$.Because $\{\mathbf{a_i}\}_{1\leqslant i\leqslant n}$ is the basis of $V \Longrightarrow \mu_i\lambda_i = 0 \xRightarrow{\lambda_i\neq 0} \mu_i = 0 \Longrightarrow \{\lambda_i\mathbf{a_i}\}_{1\leqslant i\leqslant n}$ is also a basis of V.

7.2

If the scalars $\lambda_1, ..., \lambda_n$ satisfy that$\sum_{i=1}^{n} \lambda_i b_i = 0$. Let $\mathbf{P} = (p_{ij})_{nn}$,then $b_i = \sum_{j=1}^{n} p_{ji}a_j \Longrightarrow 0 = \sum_{i=1}^{n} \lambda_i b_i = \sum_{i=1}^{n} \lambda_i \left(\sum_{j=1}^{n} p_{ji}a_j\right) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \lambda_i p_{ji}\right) a_j$.Because $\{a_i\}_{1\leqslant i\leqslant n}$ is a basis of V, $\sum_{i=1}^{n} \lambda_i p_{ji} = 0, \forall 1 \leqslant j \leqslant n \Longrightarrow \mathbf{P}(\lambda_1, \lambda_2, ..., \lambda_n)^\top = 0 \Longrightarrow \mathbf{P}^{-1}\mathbf{P}(\lambda_1, \lambda_2, ..., \lambda_n)^\top = 0 \Longrightarrow \lambda_i = 0, \forall 1 \leqslant i \leqslant n \Longrightarrow \{b_i\}_{1\leqslant i\leqslant n}$ is also a basis of V.

7.3

(a)
Notice that $\mathbf{v} = \sum_{i=1}^{n} x_i\mathbf{a_i} = \sum_{i=1}^{n} \frac{x_i}{\lambda_i}(\lambda_i\mathbf{a_i})$,so the coordinate of $\mathbf{v}$ under the basis $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, ..., \lambda_n\mathbf{a_n}\}$ is $(\frac{x_1}{\lambda_1}, \frac{x_2}{\lambda_2}, ..., \frac{x_n}{\lambda_n})$.
(b)
the coordinate of $\mathbf{w}$ under $\{\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_n}\}$ is $(1, 1, ..., 1)$;the coordinate of $\mathbf{w}$ under $\{\lambda_1\mathbf{a_1}, \lambda_2\mathbf{a_2}, ..., \lambda_n\mathbf{a_n}\}$ is $(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, ..., \frac{1}{\lambda_n})$.

7.4

(a)
If $\lambda\mathbf{c}+\mu\mathbf{b} = \mathbf{0} \Longrightarrow (-\lambda, \mu) = \mathbf{0} \Longrightarrow \lambda = \mu = 0 \Longrightarrow \{\mathbf{c}, \mathbf{b}\}$ is indeed a basis of the two-dimensional space. $\mathbf{v} = (x, y) = (-x)\mathbf{c} + y\mathbf{b}$, so the coordinate of $\mathbf{v}$ under the basis$\{\mathbf{c}, \mathbf{b}\}$ is $(-x, y)$.Because $\{\mathbf{c}, \mathbf{b}\}$ is a basis,the coordinate is unique.
(b)
$(x, y) = v = x'\mathbf{a} + y'\mathbf{b} + z'\mathbf{c} = x'(1, 0) + y'(0, 1) + z'(-1, 0) = (x' - z', y')$,so the coefficients satisfy that$x' - z' = x, y' = y$.
(c)

$(x', y', z')_{l_1} = |x'| + |y'| + |z'| \geqslant |x' - z'| + |y'| = |x| + |y|$.While the coefficients reach the minimum l_1 norm,$|x'| + |z'| = |x' - z'|$,So $x'z' \leqslant 0, x' - z' = x, y' = y$.

8 Exercise 8: Derivatives with matrices

8.1

(a)
$\lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|f(\mathbf{z})-f(\mathbf{x})-\mathbf{a}^\top(\mathbf{z}-\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} = \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{0}{\|\mathbf{z}-\mathbf{x}\|_2} = 0$, So f is differentiable and $f'(x) = \mathbf{a}^\top$.
(b)
$\lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|f(\mathbf{z})-f(\mathbf{x})-2\mathbf{x}^\top(\mathbf{z}-\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} = \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|\mathbf{z}^\top\mathbf{z}+\mathbf{x}\top\mathbf{x}-\mathbf{x}^\top\mathbf{z}-(\mathbf{x}^\top\mathbf{z})^\top\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} = \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|(\mathbf{z}^\top-\mathbf{x}^\top)(\mathbf{z}-\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2}$
$\leqslant \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|\mathbf{z}^\top-\mathbf{x}^\top\|_2\|\mathbf{z}-\mathbf{x}\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} = \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \|\mathbf{z}^\top - \mathbf{x}^\top\|_2 = 0$. So f is differentiable and $f'(x) = 2\mathbf{x}^\top$.
(c)
$\lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|f(\mathbf{z})-f(\mathbf{x})-(-2)(\mathbf{y}-\mathbf{A}\mathbf{x})^\top\mathbf{A}(\mathbf{z}-\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} = \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|(\mathbf{A}\mathbf{z}-\mathbf{A}\mathbf{x})^\top(\mathbf{A}\mathbf{z}-\mathbf{A}\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} \leqslant \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|(\mathbf{A}\mathbf{z}-\mathbf{A}\mathbf{x})^\top\|_2\|(\mathbf{A}\mathbf{z}-\mathbf{A}\mathbf{x})\|_2}{\|\mathbf{z}-\mathbf{x}\|_2} \leqslant$
$\lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|(\mathbf{A}\mathbf{z}-\mathbf{A}\mathbf{x})\|_2^2}{\|\mathbf{z}-\mathbf{x}\|_2} \leqslant \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \frac{\|\mathbf{A}\|_2^2\|\mathbf{z}-\mathbf{x}\|_2^2}{\|\mathbf{z}-\mathbf{x}\|_2} = \|\mathbf{A}\|_2^2 \lim\limits_{\mathbf{z}\to\mathbf{x},\mathbf{z}\neq\mathbf{x}} \|\mathbf{z}-\mathbf{x}\|_2 = 0$.So f is differentiable and $f'(x) = -2(\mathbf{y} - \mathbf{A}\mathbf{x})^\top\mathbf{A}$.

8.2

$\mathbf{X_0} \in \mathbb{R}^{n\times n}$be a matrix and Let $L\colon \mathbb{R}^{n\times n} \to \mathbb{R}$ be a function.We say that f is differentiable at $\mathbf{X_0}$ with derivative L if w we have

$$\lim_{\mathbf{X}\to\mathbf{X_0},\mathbf{X}\neq\mathbf{X_0}} \frac{\|f(\mathbf{X}) - f(\mathbf{X_0}) - L(\mathbf{X} - \mathbf{X_0})\|_2}{\|\mathbf{X} - \mathbf{X_0}\|_2} = 0$$

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

8.3

Let $L(\mathbf{X}) = tr(\mathbf{X}), f(\mathbf{X}) = \mathrm{tr}(\mathbf{A}^\top\mathbf{X})$

$$\begin{aligned}
&\lim_{\mathbf{X}\to\mathbf{X_0},\mathbf{X}\neq\mathbf{X_0}} \frac{\|f(\mathbf{X}) - f(\mathbf{X_0}) - L(\mathbf{X} - \mathbf{X_0})\|_2}{\|\mathbf{X} - \mathbf{X_0}\|_2} \\
&= \lim_{\mathbf{X}\to\mathbf{X_0},\mathbf{X}\neq\mathbf{X_0}} \frac{\|tr(\mathbf{X}) - tr(\mathbf{X_0}) - tr(\mathbf{X} - \mathbf{X_0})\|_2}{\|\mathbf{X} - \mathbf{X_0}\|_2} \\
&= \lim_{\mathbf{X}\to\mathbf{X_0},\mathbf{X}\neq\mathbf{X_0}} \frac{\|tr(\mathbf{X} - \mathbf{X_0} - \mathbf{X} + \mathbf{X_0})\|_2}{\|\mathbf{X} - \mathbf{X_0}\|_2} \\
&= 0
\end{aligned}$$

So, we can assert that it is differentiable,and f' is $tr(\mathbf{X})$

9 Exercise 9: Rank of Matrices

9.1

a

The rank of a matrix is defined as the dimension of its row space or column space.So,the transported matrix has the same dimension of its row space or column space.So,we can get :

$$\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top)$$

We can use SVD to prove the rest content :

$$A = U_{m\times r}\Sigma_{r\times r}V_{r\times n}^T$$

So we have:

$$\begin{aligned} A^TA &= [U_{m\times r}\Sigma_{r\times r}V_{r\times n}^T]^TU_{m\times r}\Sigma_{r\times r}V_{r\times n}^T \\ &= V_{r\times n}\Sigma_{r\times r}U_{m\times r}^TU_{m\times r}\Sigma_{r\times r}V_{r\times n}^T \\ &= V_{r\times n}\Sigma_{r\times r}^2V_{r\times n}^T \end{aligned}$$

So,we have

$$\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top\mathbf{A})$$

By the same token,we can get:

$$\begin{aligned} AA^T &= U_{m\times r}\Sigma_{r\times r}V_{r\times n}^T[U_{m\times r}\Sigma_{r\times r}V_{r\times n}^T]^T \\ &= U_{m\times r}\Sigma_{r\times r}V_{r\times n}^TV_{r\times n}\Sigma_{r\times r}U_{m\times r}^T \\ &= U_{m\times r}\Sigma_{r\times r}^2U_{m\times r}^T \end{aligned}$$

So,we have

$$\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}\mathbf{A}^\top)$$

To sum up

$$\mathbf{rank}(\mathbf{A}) = \mathbf{rank}(\mathbf{A}^\top) = \mathbf{rank}(\mathbf{A}^\top\mathbf{A}) = \mathbf{rank}(\mathbf{A}\mathbf{A}^\top);$$

b

Every column of AB is a combination of the columns of A (matrix multiplication),so

$$\text{rank}(\mathbf{AB}) \le \text{rank}(\mathbf{A})$$

Every row of AB is a combination of the rows of B (matrix multiplication)
When we consider row rank = column rank,we can get

$$\text{Rank}(AB) \le \min(\text{Rank}(A), \text{Rank}(B))$$

Let,

$$A = (\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix})$$

$$B = (\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix})$$

So,

$$AB = (\begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix})$$

we get rank(AB)=2=rank(A).

9.2

a

Let us consider this: Definition 29. Let $\mathbf{A} \in \mathbb{R}^m \times n$. The range of $\mathbf{A}$, denoted $\mathcal{R}(\mathbf{A})$, is the set of all vectors in $\mathbb{R}^m$ that can be written as linear combinations of the columns of $\mathbf{A}$, i.e.,

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{Ax} \in \mathbb{R}^m \mid \mathbf{x} \in \mathbb{R}^n\}.$$

So we can say c(A) = R(A),then let us see the definition of rank:

2. *The dimension of* $\mathcal{R}(\mathbf{A})$ *is the rank of* $\mathbf{A}$, *i.e.,* $\dim \mathcal{R}(\mathbf{A}) = \operatorname{rank} \mathbf{A}$.

So we can easily get the conclusion:

$$\operatorname{rank}(\mathbf{A}) = \dim(\mathcal{C}(\mathbf{A}))$$

b

$$A_{m\times n}X = 0$$

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} X = 0$$

Gaussian Elimination

$$\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \cdots & & \cdots & b_{rn} \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 0 \end{pmatrix} X = 0$$

So w can easily find that :

$$\dim(\mathcal{C}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A})) = n$$

b

Let us take this into 1.b,then we just need to prove

$$rank(B) - rank(A) \leq \dim(\mathcal{C}(\mathbf{B}) \cap \mathcal{N}(\mathbf{A})).$$

$$\dim(\mathcal{C}(\mathbf{B})) - (n - \dim(\mathcal{N}(\mathbf{A}))) \leq \dim(\mathcal{C}(\mathbf{B}) \cap \mathcal{N}(\mathbf{A}))$$

$$\dim(\mathcal{C}(\mathbf{B})) + \dim(\mathcal{N}(\mathbf{A})) \leq \dim(\mathcal{C}(\mathbf{B}) \cap \mathcal{N}(\mathbf{A})) + n$$

We know:

$$\dim(\mathcal{C}(\mathbf{B}))+\dim(\mathcal{N}(\mathbf{A})) = \dim(\mathcal{C}(\mathbf{B})+\mathcal{N}(\mathbf{A}))+\dim(\mathcal{C}(\mathbf{B})\cap\mathcal{N}(\mathbf{A})) \leq \dim(\mathcal{C}(\mathbf{B})\cap\mathcal{N}(\mathbf{A}))+n$$

So we have :

$$\operatorname{rank}(\mathbf{AB}) \leq \operatorname{rank}(\mathbf{A})$$

10 Exercise 10: Properties of Eigenvalues and Singular Values

10.1

$$\lambda_{\max}(\mathbf{A}) = \sup_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^\top \mathbf{A}\mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

Thinking that A is a symmetric matrix,we can get :

$$\mathbf{A} = \mathbf{P}^\top \Sigma \mathbf{P}$$

Take this into λ:

$$\begin{aligned} &\frac{\mathbf{x}^\top \mathbf{P}^\top \Sigma \mathbf{P}\mathbf{x}}{\mathbf{x}^\top \mathbf{x}} \\ &= \frac{(\mathbf{P}\mathbf{x})^\top \Sigma (\mathbf{P}\mathbf{x})}{\mathbf{x}^\top \mathbf{P}^\top \mathbf{P}\mathbf{x}} \\ &= \frac{(\mathbf{x}\mathbf{P})^\top \Sigma (\mathbf{P}\mathbf{x})}{(\mathbf{P}\mathbf{x})^\top (\mathbf{x}\mathbf{P})} \end{aligned}$$

Let $\mathbf{Px}$ be $\mathbf{y} \in \mathbb{R}^n, \mathbf{y} \neq \mathbf{0}$

We have:

$$\lambda_{\max}(\mathbf{A}) = \sup_{\mathbf{y}\in\mathbb{R}^n, \mathbf{y}\neq\mathbf{0}} \frac{\mathbf{y}^\top \boldsymbol{\Sigma}\mathbf{y}}{\mathbf{y}^\top \mathbf{y}}$$

thinking that

$$\frac{\sum_{j=1}^n \lambda_j y_i^2}{\sum_{j=1}^n {y_i}^2} \leq \max_{1\leq j\leq n} \lambda_j$$

So we have

$$\lambda_{\max}(\mathbf{A}) = \sup_{\mathbf{y}\in\mathbb{R}^n, \mathbf{y}\neq\mathbf{0}} \frac{\mathbf{y}^\top \mathbf{\Sigma} \mathbf{y}}{\mathbf{y}^\top \mathbf{y}} = \max_{1\leq j\leq n} \lambda_j$$

And this is the same:

$$\lambda_{\min}(\mathbf{A}) = \inf_{\mathbf{x}\in\mathbb{R}^n, \mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}.$$