CIS4930 - Math for Machine Learning

Homework 5

Fernando Sckaff

March 26, 2023

Question 1

Let K be the set of all numbers which can be written in the form $a+bi$ where $i = \sqrt{-1}$ and a, b are rational numbers. Show that K is a field.

Solution:
K is a field if it satisfies following conditions:

1. If x, y are elements of K, then $x + y$ and xy are also elements of K
2. If $x \in K$, then $-x$ is also an element of K
3. If $x \neq 0$, then x^{-1} is an element of K
4. The elements 0 and 1 are elements K.

1.
$x = a + bi \in K$
$y = c + di \in K$
$a, b, c, d \in \mathbb{Q}$

$$\begin{aligned} x + y &= (a + bi) + (c + di) \\ &= a + bi + c + di \\ &= (a + c) + (b + d)i \end{aligned}$$

Since $a, b, c, d \in \mathbb{Q}$, then $(a + c) \in \mathbb{Q}$ and $(b + d) \in \mathbb{Q}$. Therefore, addition property holds.
Also,

$$\begin{aligned} xy &= (a + bi)(c + di) \\ &= ac + adi + cbi + bdi^2 \\ &= (ac - bd) + (ad + cb)i \end{aligned}$$

Since $a, b, c, d \in \mathbb{Q}$, then $(ac - bd) \in \mathbb{Q}$ and $(ad + cb) \in \mathbb{Q}$. Therefore, multiplication property holds.
2.
Let $x = a + bi$

$$\begin{aligned} -x &= -(a + bi) \\ &= (-a) + (-b)i \end{aligned}$$

Since $a, b \in \mathbb{Q}$, then $-a, -b$ also $\in \mathbb{Q}$. Therefore, negative property holds.
3.
Let $x = a + bi$ and $t = a^2 + b^2 \in \mathbb{Q}$

$$\begin{aligned} x^{-1} &= \frac{1}{a + bi} \\ &= \frac{1}{a + bi} \cdot \frac{a - bi}{a - bi} \\ &= \frac{a - bi}{a^2 - (bi)^2} \\ &= \frac{a - bi}{a^2 + b^2} \\ &= \frac{a - bi}{t} \\ &= (a/t) + (-b/t)i \end{aligned}$$

Since $a, b, t \in \mathbb{Q}$, then $(a/t), (-b/t) \in \mathbb{Q}$. Therefore, inverse property holds.

4.
Let $x = a + bi$, $a = 0$, and $b = 0$.

$$x = 0 + 0i = 0$$

Now, let $a = 1$ and $b = 0$.

$$x = 1 + 0i = 1$$

Therefore, 0 and 1 are in the field. 😄

Question 2

If U and W are subspaces of a vector space V, show that
1. $U \cap W$ (the set of elements that lie both in U and W) and
2. $U + W$ (the set of all elements $u + w$ with $u \in U$ and $w \in W$)
are subspaces of V.

Solution:
W is a subspace of V if it satisfies the following conditions:

- If v, w are elements of W, their sum $v + w$ is also an element of W
- If v is an element of W and $c \in K$, then cv is an element of W
- The element O of V is also an element of W

1.
Let

- $T \equiv U \cap W$
- $u, w \in T$

Then $u + w$ is $((u_1 + w_1), \ldots, (u_n + w_n))$.
First

$$u + w \in U \text{ since } u \in U \text{ and } w \in W.$$

Similarly, $u + w \in W$. Therefore, $u + w \in T$
Second

$$cu \in T \text{ since } cu \in U \text{ and } cw \in W$$

Third

$$O \in T \text{ since } O \in U \text{ and } O \in W$$

2.
Let $u, w \in U + W$. Then $u + w \in U + W$
First

$$u + w \in U + W \text{ since } u \in U \text{ and } w \in W$$

Second
Since $u \in U, U \in U + W (O \in W),$ and $c \in W$

$$cu \in U + W$$

Third

$$O \in U \;\; \& \;\; O \in W \therefore O \in U + W$$

Question 3

Show that $\|A\|_1$ which is equal to $\max_j \sum_{i=1}^{m} |a_{ij}|$ and $\|A\|_\infty$ which is equal to $\max_i \sum_{j=1}^{n} |a_{ij}|$ satisfy the following properties.

1. $\|A\| > 0$ if $A \neq O$ (the all zero matrix).
2. $\|\gamma A\| = |\gamma| \cdot \|A\|$ for γ.
3. $\|A + B\| \leqslant \|A\| + \|B\|$,
4. $\|AB\| \leqslant \|A\| \cdot \|B\|$.
5. $\|Ax\| \leqslant \|A\| \cdot \|x\|$ for any vector x (and for the same choice of vector norm).

Solution: Solution question something

1. First, assume $A \neq O$, and let's say $\|A\|_1 < 0$.

$$
\begin{aligned}
\|A\|_1 &< 0 \\
\max_j \sum_{i=1}^{m} |a_{ij}| &< 0 \\
\max_j (|a_{1j}| + |a_{2j}| + \ldots + |a_{ij}|) &< 0
\end{aligned}
$$

A sum of non-zero, absolute values cannot be negative. Therefore, proof by contradiction. 😄
Likewise,

$$
\begin{aligned}
\|A\|_\infty &< 0 \quad ? \\
\max_i \sum_{j=1}^{m} |a_{ij}| &< 0 \\
\max_i (|a_{i1}| + |a_{i2}| + \ldots + |a_{ij}|) &< 0 \\
\Rightarrow \max_i (|a_{i1}| + |a_{i2}| + \ldots + |a_{ij}|) &\not< 0 \quad 😄
\end{aligned}
$$

2.

$$
\begin{aligned}
\|\gamma A\|_1 &= \max_j \sum_{i=1}^{m} |\gamma a_{ij}| \\
&= \max_j |\gamma| \sum_{i=1}^{m} |a_{ij}| \\
&= |\gamma| \max_j \sum_{i=1}^{m} |a_{ij}| \\
&= |\gamma| \cdot \|A\|_1 \qquad 😄
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\|\gamma A\|_\infty &= \max_i \sum_{j=1}^{m} |\gamma a_{ij}| \\
&= \max_i |\gamma| \; \sum_{j=1}^{m} |a_{ij}| \\
&= |\gamma| \max_i \sum_{j=1}^{m} |a_{ij}| \\
&= |\gamma| \cdot \|A\|_\infty && \text{😄}
\end{aligned}
$$

3. Remember, $|A+B| \leqslant |A| + |B|$.

$$
\begin{aligned}
\|A+B\|_1 &= \max_j \sum_{i=1}^{m} |a_{ij} + b_{ij}| \\
\max_j \sum_{i=1}^{m} |a_{ij} + b_{ij}| &\leqslant \max_j \sum_{i=1}^{m} |a_{ij}| + \max_j \sum_{i=1}^{m} |b_{ij}| \\
&\leqslant \|A\|_1 + \|B\|_1 && \text{😄}
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\|A+B\|_\infty &= \max_i \sum_{j=1}^{m} |a_{ij} + b_{ij}| \\
\max_i \sum_{j=1}^{m} |a_{ij} + b_{ij}| &\leqslant \max_i \sum_{j=1}^{m} |a_{ij}| + \max_i \sum_{j=1}^{m} |b_{ij}| \\
&\leqslant \|A\|_\infty + \|B\|_\infty && \text{😄}
\end{aligned}
$$

4. Remember, $|AB| \leqslant |A| \cdot |B|$.

$$
\begin{aligned}
\|AB\|_1 &= \max_j \sum_{i=1}^{m} |a_{ij} b_{ij}| \\
\max_j \sum_{i=1}^{m} |a_{ij} b_{ij}| &\leqslant \max_j \sum_{i=1}^{m} |a_{ij}| \cdot \max_j \sum_{i=1}^{m} |b_{ij}| \\
&\leqslant \|A\|_1 \cdot \|B\|_1 \quad \text{😄}
\end{aligned}
$$

Likewise,

$$
\begin{aligned}
\|AB\|_\infty &= \max_i \sum_{j=1}^{m} |a_{ij} b_{ij}| \\
\max_i \sum_{j=1}^{m} |a_{ij} b_{ij}| &\leqslant \max_i \sum_{j=1}^{m} |a_{ij}| \cdot \max_i \sum_{j=1}^{m} |b_{ij}| \\
&\leqslant \|A\|_\infty \cdot \|B\|_\infty \quad \text{😄}
\end{aligned}
$$

5.

$$\|Ax\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}x_j|$$

$$\|x\|_1 = \sum_{i=1}^{m} |x_i|$$

Assume $\|Ax\|_1 \geqslant \|A\|_1 \cdot \|x\|_1$:

$$\|Ax\|_1 \geqslant \|A\|_1 \cdot \|x\|_1$$

$$\max_j \sum_{i=1}^{m} |a_{ij}x_j| \geqslant \max_j \sum_{i=1}^{m} |a_{ij}| \cdot \sum_{i=1}^{m} |x_i|$$

$$\max_j \sum_{i=1}^{m} |a_{ij}x_j| \geqslant \max_j \sum_{i=1}^{m} |a_{ij}| \cdot \sum_{i=1}^{m} |x_i|$$

$$\max_j \sum_{i=1}^{m} |a_{ij}x_j| \geqslant \max_j \sum_{i=1}^{m} (|a_{ij}| \cdot |x_i|)$$

As we know, $|a_{ij}x_i| \leqslant |a_{ij}| \cdot |x_i|$. Therefore, a contradiction. 😄
Also, assume $\|Ax\|_\infty \geqslant \|A\|_\infty \cdot \|x\|_\infty$:

$$\|Ax\|_\infty \geqslant \|A\|_\infty \cdot \|x\|_\infty$$

$$\max_i \sum_{j=1}^{m} |a_{ij}x_j| \geqslant \max_i \sum_{j=1}^{m} |a_{ij}| \cdot \max_i |x_i|$$

$$\max_i \sum_{j=1}^{m} |a_{ij}x_j| \geqslant \max_i \sum_{j=1}^{m} |a_{ij}| \cdot |x_i| \quad 😄$$

As we know, $|a_{ij}x_i| \leqslant |a_{ij}| \cdot |x_i|$. Therefore, a contradiction. 😄

Question 4

(a) Show that the quadratic form $x^T Bx$ annihilates the skew symmetric portion of B (where we can write $B = \frac{B+B^T}{2} + \frac{B-B^T}{2}$).
(b) Is the matrix $B \equiv A^T A$ non-negative definite?
(Hint: For a non-negative definite matrix B, the quadratic form $x^T Bx \geqslant 0$ for all $x \neq O$.) Explain.

Solution: (a) We have $B = \frac{B+B^T}{2} + \frac{B-B^T}{2}$, where the first term is symmetric and the second term is skew-symmetric. Then we have:

$$\begin{aligned} x^T Bx &= x^T \left(\frac{B+B^T}{2} + \frac{B-B^T}{2} \right) x \\ &= x^T \frac{B+B^T}{2} x + x^T \frac{B-B^T}{2} x \end{aligned}$$

Be mindful, a skew-symmetric matrix is a square matrix whose transpose equals its negative. ($B^T = -B$). Assume B is skew-symmetric:

$$
\begin{aligned}
x^T B x &= x^T \frac{B + B^T}{2} x + x^T \frac{B - B^T}{2} x \\
&= x^T \frac{B - B}{2} x + x^T \frac{B + B}{2} x \\
&= x^T \cdot 0 \cdot x + x^T \frac{2B}{2} x \\
&= x^T B x
\end{aligned}
$$

Therefore, $B = \frac{B+B^T}{2} + \frac{B-B^T}{2}$ does eliminate the skew-symmetric part of B.
(b) Let x be a column vector. Then we have:

$$x^T B x = x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2$$

Since the norm of a vector is non-negative, we have $x^T Bx \geqslant 0$ for all $x \neq O$. Therefore, the matrix $B = A^T A$ is non-negative definite.

Question 5

Rewrite the least-square orthogonality principle($A^T(b - A\hat{x} = O)$ where $\hat{x}$ is the least-squares solution for the problem $\min_x \|b - Ax\|_2^2$ in the form $C\hat{x} = Db$ where C and D are written in terms of the SVD of A (which can be written as $A = USV^T$). C should not contain U and D should not contain V. Note that $U^T U = I$ and $V^T V = I$. (Note: You may use the reduced form of the SVD if you wish.)

Solution: Starting with the least-square orthogonality principle: $A^T(b - Ax) = 0$ Using the SVD of A ($A = USV^T$), we have:

$$
\begin{aligned}
A^T(b - Ax) &= 0 \\
A^T b - A^T A x &= 0 \\
A^T A x &= A^T b \\
V S^T U^T U S V^T x &= V S^T U^T b
\end{aligned}
$$

Simplifying using $U^T U = I$ and $V^T V = I$, we get:

$$S V^T x = U^T b$$

Multiplying both sides by S^{-1}, we have:

$$V^T x = S^{-1} U^T b$$

Multiplying both sides by V, we have:

$$x = V S^{-1} U^T b$$

Thus, we have expressed the least-squares solution x in terms of the SVD of A as $x = VS^{-1} U^T b$.
We can write C and D in terms of the SVD of A as follows:

$$C = V \text{ (since } V^T V = I) \text{ and } D = S^{-1} U^T$$

Therefore, the least-squares solution can be expressed as: $Cx = Db$ where $C = V$ and $D = S^{-1} U^T$, and $A = USV^T$.

Question 6

Set up the Lagrangian for the following objective function:

- $f(x) = -\sum_{i=1}^N T_i x_i + \sum_{i=1}^N x_i \log x_i$ subject to the constraint $\sum_{i=1}^N x_i = 1$ where T is a set of real numbers. Solve for the Lagrange parameter and then find the solution for x (called the **softmax** nonlinearity without the Lagrange parameter).

Solution: The objective function is:

$$f(x) = -\sum_{i=1}^{N} T_i x_i + \sum_{i=1}^{N} x_i \log x_i$$

subject to the constraint:

$$\sum_{i=1}^{N} x_i = 1$$

To set up the Lagrangian, we introduce a Lagrange multiplier λ:

$$\mathcal{L}(x, \lambda) = f(x) + \lambda \left(\sum_{i=1}^{N} x_i - 1 \right)$$

where $\mathcal{L}$ is the Lagrangian.

Now we compute the partial derivatives of $\mathcal{L}$ with respect to x_i and λ, and set them equal to zero:

$$\frac{\partial \mathcal{L}}{\partial x_i} = -T_i + \log x_i + 1 + \lambda = 0 \quad \text{for } i = 1, \ldots, N \tag{1}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^{N} x_i - 1 = 0 \tag{2}$$

Solving the first equation for x_i gives:

$$x_i = e^{T_i - 1 - \lambda}$$

Substituting this expression into the second equation, we have:

$$\sum_{i=1}^{N} e^{T_i - 1 - \lambda} = 1$$

$$\log(\sum_{i=1}^{N} e^{T_i - 1 - \lambda}) = \log 1$$

$$\log(\sum_{i=1}^{N} e^{T_i - 1} \cdot e^{-\lambda}) = 0$$

$$\log \sum_{i=1}^{N} e^{T_i - 1} + \log \sum_{i=1}^{N} e^{-\lambda} = 0$$

$$\log \sum_{i=1}^{N} e^{T_i - 1} - \lambda = 0$$

which implies:

$$\lambda = \log \left(\sum_{i=1}^{N} e^{T_i - 1} \right)$$

Therefore, by substituting λ inside $-T_i + \log x_i + 1 + \lambda = 0$, the solution for x_i is:

$$\begin{aligned}
-T_i + \log x_i + 1 + \lambda &= 0 \\
\log x_i &= T_i - \lambda - 1 \\
\log x_i &= T_i - \log(\sum_{i=1}^{N} e^{T_i - 1}) - 1 \\
x_i &= e^{T_i - \log(\sum_{i=1}^{N} e^{T_i - 1}) - 1} \\
x_i &= e^{T_i} \cdot e^{-\log(\sum_{i=1}^{N} e^{T_i - 1})} \cdot e^{-1} \\
x_i &= \frac{e^{T_i - 1}}{\sum_{j=1}^{N} e^{T_j - 1}}
\end{aligned}$$

This is known as the softmax nonlinearity without the Lagrange parameter. 😊

F.S.

Note:-

- Question 1: Correct
- Question 2: $(AB)_{ij} \neq a_{ij}b_{ij}$ which is what you are trying to do. $(AB)_{ij} = \sum_k a_{ik}b_{kj}$ and to make your approach work, you have to use this matrix multiplication formula and then bound the sum. Take a look at my revised solution which expands on this.
- Question 3: Correct
- Question 4: Correct
- Question 5: Correct. OBS: Could have moved S^{-1} to the other side but we'll let this slide.
- Question 6: Correct. OBS: Could have eliminated the -1 from the numerator AND denominator exponents. You would then have the standard softmax nonlinearity $x_i = \exp T_i / \sum_k \exp T_k$.