Solutions to HW1

1. Consider the vector space $\mathbb{R}^n$.

   (a) Check that $\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n} |x_i|$ is indeed a norm on $\mathbb{R}^n$.

   (b) Prove that: for any $\boldsymbol{x} \in \mathbb{R}^n$,
   $$\|\boldsymbol{x}\|_\infty = \lim_{p\to\infty} \|\boldsymbol{x}\|_p.$$

   (c) Prove the equivalence
   $$\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_1 \le n\|\boldsymbol{x}\|_\infty, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Solution: (a) 1) $\|x\|_\infty = \max\limits_{1\le i\le n} |x_i| \geq 0$

If $\|x\|_\infty = 0$, then $|x_i| = 0$ for all $1\le i\le n$.

That is, $x_i = 0$, $1\le i\le n$ and $\vec{x} = \vec{0}$

Hence $\|x\|_\infty = 0$ if and only if $\vec{x} = \vec{0}$.

2) For any $\alpha \in \mathbb{R}$ and $\vec{x} \in \mathbb{R}^n$,

$$\begin{aligned}
\|\alpha\vec{x}\|_\infty &= \max_{1\le i\le n} |\alpha x_i| \\
&= \max_{1\le i\le n} |\alpha|\,|x_i| \\
&= |\alpha| \max_{1\le i\le n} |x_i| \\
&= |\alpha|\, \|x\|_\infty
\end{aligned}$$

3) For any $\vec{x}, \vec{y} \in \mathbb{R}^n$

$$\begin{aligned}
\|\vec{x}+\vec{y}\|_\infty &= \max_{1\le i\le n} |x_i + y_i| \\
&\le \max_{1\le i\le n} (|x_i| + |y_i|) \\
&\le \max_{1\le i\le n} |x_i| + \max_{1\le i\le n} |y_i| \\
&= \|\vec{x}\|_\infty + \|\vec{y}\|_\infty
\end{aligned}$$

Based on 1), 2) and 3), $\|\vec{x}\|_\infty = \max\limits_{1\le i\le n} |x_i|$ is a norm.

b) Assume $p>1$. Then

$$\|x\|_\infty = \max_i |x_i| = \left(\max_i |x_i|^p\right)^{\frac{1}{p}} \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} = \|x\|_p$$

$$\le \left(\sum_{i=1}^{n} \max_i |x_i|^p\right)^{\frac{1}{p}} \le \left(n \max_i |x_i|^p\right)^{\frac{1}{p}}$$

$$= n^{\frac{1}{p}} \left(\max_i |x_i|\right)^{p \times \frac{1}{p}} = n^{\frac{1}{p}} \|x\|_\infty$$

That is, $\|x\|_\infty \le \|x\|_p \le n^{\frac{1}{p}} \|x\|_\infty$

Since $\lim_{p\to\infty} n^{\frac{1}{p}} = 1$, we have

$$\lim_{p\to\infty} n^{\frac{1}{p}} \|x\|_\infty = \|x\|_\infty .$$

By Sandwich theorem, we have

$$\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$$

c) $\|x\|_\infty = \max_i |x_i| \le \sum_i |x_i| = \|x\|_1$

$\le n \max_i |x_i| = n \|x\|_\infty$

Hence $\|x\|_\infty \le \|x\|_1 \le n \|x\|_\infty$, $\forall x \in \mathbb{R}^n$

2. For any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we have defined

$$\|\boldsymbol{A}\|_2 = \sup_{\boldsymbol{x}\in\mathbb{R}^n,\ \boldsymbol{x}\neq\boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2}.$$

(a) Prove that

$$\|\boldsymbol{A}\|_2 = \max_{\boldsymbol{x}\in\mathbb{R}^n,\ \|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2$$

(b) Prove that $\|\cdot\|_2$ is a norm on $\mathbb{R}^{m\times n}$.

(c) Prove that $\|\boldsymbol{A}\boldsymbol{x}\|_2 \le \|\boldsymbol{A}\|_2\|\boldsymbol{x}\|_2$ for any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{R}^n$.

(d) Prove that $\|\boldsymbol{A}\boldsymbol{B}\|_2 \le \|\boldsymbol{A}\|_2\|\boldsymbol{B}\|_2$ for all $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times p}$.

Solution: (a) Note that for any $\vec{x} \in \mathbb{R}^n$ and $\vec{x} \neq 0$, we have

$$\left\| \frac{\vec{x}}{\|\vec{x}\|} \right\| = \frac{\|\vec{x}\|}{\|\vec{x}\|} = 1 .$$

Hence $\|A\|_2 = \sup_{\vec{x}\in\mathbb{R}^n, \vec{x}\neq 0} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} = \sup_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\vec{x}\|_2$

Since $\|A\vec{x}\|_2$ is a continuous function, the suprimum can be achieved. Hence $\|A\|_2 = \sup_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\vec{x}\|_2 = \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\vec{x}\|_2$

(b) 1) Since $\|A\vec{x}\|_2 \geq 0$, we have

$$\|A\|_2 = \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|_2=1} \|A\vec{x}\|_2 \geq 0.$$

If $\|A\|_2 = 0$, then $\|A\vec{x}\|_2 = 0$ for all $\vec{x} \in \mathbb{R}^n$, $\|\vec{x}\|_2 = 1$

Let $A = [\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_n]$. If we choose $\vec{x} = \vec{e}_1$, then

$\|A\vec{e}_1\|_2 = \|\vec{a}_1\|_2 = 0$ which implies $\vec{a}_1 = \vec{0}$.

Similarly, we can get $\vec{a}_2 = \vec{a}_3 = \cdots = \vec{a}_n = \vec{0}$.

Therefore, $\|A\|_2 = 0$ if and only if $A = \vec{0}$

2) For any $\alpha \in \mathbb{R}$, we have

$$\begin{aligned}
\|\alpha A\|_2 &= \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|\alpha A\vec{x}\|_2 \\
&= \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} |\alpha|\, \|A\vec{x}\|_2 \\
&= |\alpha| \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\vec{x}\|_2 \\
&= |\alpha|\, \|A\|_2
\end{aligned}$$

3) For any $A, B \in \mathbb{R}^{m\times n}$, we have

$$\begin{aligned}
\|A+B\|_2 &= \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|(A+B)\vec{x}\|_2 \\
&\leq \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} (\|A\vec{x}\|_2 + \|B\vec{x}\|_2) \\
&\leq \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\vec{x}\|_2 + \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|B\vec{x}\|_2 \\
&= \|A\|_2 + \|B\|_2
\end{aligned}$$

Based on 1), 2) and 3), $\|A\|_2 = \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\vec{x}\|_2$ is a norm.

(c) Note that $\|A\|_2 = \sup_{\vec{x}\in\mathbb{R}^n, \vec{x}\neq\vec{0}} \frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2}$. We have

$$\frac{\|A\vec{x}\|_2}{\|\vec{x}\|_2} \leq \|A\|_2 \quad \text{for any } \vec{x} \neq \vec{0}, \vec{x} \in \mathbb{R}^n.$$

That is, $\|A\vec{x}\|_2 \le \|A\|_2 \|x\|_2$.

d) $\|AB\|_2 = \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|AB\vec{x}\|_2 \le \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|A\|_2 \|B\vec{x}\|_2$

$= \|A\|_2 \max_{\vec{x}\in\mathbb{R}^n, \|\vec{x}\|=1} \|B\vec{x}\|_2$

$= \|A\|_2 \cdot \|B\|_2$

3. Let $a_1, a_2, \ldots, a_m$ be m given real numbers. Prove that a median of $a_1, a_2, \ldots, a_m$ minimizes

$$|a_1 - b| + |a_2 - b| + \ldots + |a_m - b|$$

over all $b \in \mathbb{R}$.

Solution: Without of genelarity, we assume $a_1 < a_2 < \cdots a_m$.

1) When m is odd,

Then $|a_1 - b| + |a_m - b|$ is minimized if $a_1 < b < a_m$

$|a_2 - b| + |a_{m-1} - b|$ is minimized if $a_2 < b < a_{m-1}$

$\vdots$

$|a_{\frac{m-1}{2}} - b| + |a_{\frac{m+3}{2}} - b|$ is minimized if $a_{\frac{m-1}{2}} < b < a_{\frac{m+3}{2}}$

$|a_{\frac{m+1}{2}} - b|$ is minimized if $b = a_{\frac{m+1}{2}}$

That is, if we choose $b = a_{\frac{m+1}{2}}$ (the median of $a_1, \cdots, a_m$)

$|a_1 - b| + |a_2 - b| + \cdots + |a_m - b|$ is minimized.

2) When m is even

Then $|a_1 - b| + |a_m - b|$ is minimized if $a_1 < b < a_m$

$|a_2 - b| + |a_{m-1} - b|$ is minimized if $a_2 < b < a_{m-1}$

$\vdots$

$|a_{\frac{m}{2}} - b| + |a_{\frac{m}{2}+1} - b|$ is minimized if $a_{\frac{m}{2}} < b < a_{\frac{m}{2}+1}$

In particular, a median of $a_1, \cdots, a_m$ minimize

$|a_1 - b| + |a_2 - b| \cdots + |a_m - b|$.

Combining 1), 2), a median of $a_1, \cdots, a_m$ minimize

$|a_1 - b| + |a_2 - b| + \cdots + |a_m - b|$

4. Suppose that the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ in $\mathbb{R}^n$ are clustered using the K-means algorithm, with group representatives $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$.

   (a) Suppose the original vectors $\boldsymbol{x}_i$ are nonnegative, i.e., their entries are nonnegative. Explain why the representatives $\boldsymbol{z}_j$ output by the K-means algorithm are also nonnegative.

   (b) Suppose the original vectors $\boldsymbol{x}_i$ represent proportions, i.e., their entries are nonnegative and sum to one. (This is the case when $\boldsymbol{x}_i$ are word count histograms, for example.) Explain why the representatives $\boldsymbol{z}_j$ output by the K-means algorithm are also represent proportions (i.e., their entries are nonnegative and sum to one).

   (c) Suppose the original vectors $\boldsymbol{x}_i$ are Boolean, i.e., their entries are either 0 or 1. Give an interpretation of $(\boldsymbol{z}_j)_i$, the i-th entry of the j group representative.

Solution: (a) z_j is the average of x_k in the group j,

i.e. $$z_j = \frac{1}{|G_j|} \sum_{k \in G_j} x_k ,$$

where $|G_j|$ is the number of element in G_j.

Hence if x_k is non-negative, then z_j is also non-negative

(b) If the entries of x_k are nonnegative and sum to one, i.e. $1^T x_k = 1$, for all k

then
$$1^T z_j = \frac{1}{|G_j|} \sum_{k \in G_j} 1^T x_k = \frac{|G_j|}{|G_j|} = 1$$

Hence the entries of z_j are nonnegative and sums to one.

(c) $(z_j)_i$ represent the proportion of the vectors in jth group that have one in ith entry. If $(z_j)_i = 1$, all vectors in the group j have one at ith entry.

If $(z_j)_i = 0$, no vectors in the group j have one at ith entry.
If $(z_j)_i = 0.5$, half of the vectors in the group j have one at ith entry.