MATH3332

Data Analytic Tools

Larry128

A summary notes for revision

Fall 2024-2025

Contents

1 Vector Space, metric, limits/ convergence **2**
1.1 Vector Spaces . . . 2
1.2 Metric on vector space . . . 4
1.3 Matrix Vector Norm . . . 6
1.4 Matrix Norm . . . 6
1.5 Other norms . . . 8

2 Case Study: K-means clustering, K-medians clustering **9**
2.1 Clustering . . . 9
2.2 K-means clustering . . . 9
2.3 K-medians Clustering . . . 11
2.4 Comparison of K-means and K-medians . . . 11

3 Limit and Convergence on vector space **12**
3.1 Limit and convergence on a normed vector space . . . 12

1 Vector Space, metric, limits/ convergence

1.1 Vector Spaces

A vector space (linear space) over $\mathbb{R}$ (in real domain) is a set V together with functions:

1. Vector addition

$$\begin{aligned} &(V,V)\mapsto V \\ \equiv (\mathbf{x},\mathbf{y})\in(V,V) &\implies \mathbf{x}+\mathbf{y} \end{aligned}$$

2. Scalar multiplication

$$\begin{aligned} &(\mathbb{R},V)\mapsto V \\ \equiv \forall\alpha\in\mathbb{R},\mathbf{x}\in V,&\alpha\mathbf{x}\in V \end{aligned}$$

These two functions should satisfy the following 8 properties. $\forall\mathbf{x},\mathbf{y},\mathbf{z}\in V$

1. Associativity of addition

$$\mathbf{x}+\mathbf{y}+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})=(\mathbf{x}+\mathbf{y})+\mathbf{z}$$

2. Commutativity of addition

$$\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$$

3. Zero vector

$$\exists\mathbf{0}\text{ s.t. }\mathbf{x}+\mathbf{0}=\mathbf{0}+\mathbf{x}$$

4. Negative vector

$$\forall\mathbf{x}\in V,\exists-\mathbf{x}\in V,\text{ s.t. }\mathbf{x}+(-\mathbf{x})=(-\mathbf{x})+\mathbf{x}=\mathbf{0}$$

5.

$$1\dot{\mathbf{x}}=\mathbf{x}$$

6.

$$\forall\alpha,\beta\in\mathbb{R},\alpha(\beta\mathbf{x})=(\beta\alpha)\mathbf{x}$$

7.

$$\forall\alpha,\beta\in\mathbb{R},(\alpha+\beta)\mathbf{x}=\alpha\mathbf{x}+\beta\mathbf{x}$$

8.

$$\forall\alpha\in\mathbb{R},\alpha(\mathbf{x}+\mathbf{y})=\alpha\mathbf{x}+\alpha\mathbf{y}$$

Remarks:

1. We can define vector space over the complex domain $\mathbb{C}$ similarly.
2. We will assume vector space in the real domain by default. Vector space over complex domain is used very rarely.

Examples of vector spaces

1. $\mathbb{R}$ with standard addition of real numbers and the standard multiplication of real numbers is a vector space.

2. $\mathbb{R}^n$ n-dimensional Euclidean space with addition and multiplication defined as follows
   Addition:
   $$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_1 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$
   Multiplication:
   $$\alpha \in \mathbb{R}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$$
   Zero:
   $$\mathbf{0} = \begin{bmatrix} 0 & 0 & \dots & 0 \end{bmatrix}^T$$
   Then, $\mathbb{R}^n$ is a vector space since it is closed in addition and scalar multiplication, and $\mathbf{0} \in \mathbb{R}^n$.

3. All real $m \times n$ matrices $\mathbb{R}^{m \times n}$ with addition and multiplication defined as:
   Addition:
   $$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & & \\ y_{m1} & y_{m2} & \dots & y_{mn} \end{bmatrix}, X + Y = (x_{ij} + y_{ij})_{m \times n}$$
   Multiplication:
   $$\alpha \in \mathbb{R}, X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, \alpha X = (\alpha x_{ij})_{m \times n}$$
   Zero:
   $$\mathbf{0} = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \dots & 0 \end{bmatrix}$$
   Then $\mathbb{R}^{m \times n}$ is a vector space. Remarks:
   (a) This vector space is same as $\mathbb{R}^{mn}$ by vectorization.
   (b) Vectorization $\mathbb{R}^{m \times n} \mapsto \mathbb{R}^{mn}$
   $$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \to \begin{bmatrix} x_{11} \\ \vdots \\ x_{m1} \\ x_{12} \\ \vdots \\ x_{m2} \\ \vdots \\ x_{1n} \\ \vdots \\ x_{mn} \end{bmatrix}$$

4. All real 3-array of size $m \times n \times n \times l$ $\mathbb{R}^{m\times n\times l}$ with addition and multiplication defined as
Addition:
$$X = (x_{ijk})_{i,j,k}, Y = (y_{ijk})_{i,j,k} \in \mathbb{R}^{m\times n\times l}, X + Y = (x_{ijk} + y_{ijk})_{i,j,k} \in \mathbb{R}^{m\times n\times l}$$
Multiplication:
$$\alpha \in \mathbb{R}, X = (x_{ijk})_{i,j,k} \in \mathbb{R}^{m\times n\times l}, \alpha X = (\alpha x_{ijk})_{i,j,k}$$

5. (Function space). The set of all continuous functions on [a, b], denoted by
$$C[a, b] := \{f : f \text{ is a continous function on } [a, b]\}$$
with addition defined as
$$\forall f, g \in C[a, b], (f + g)(t) = f(t) + g(t), \forall t \in [a, b]$$
and multiplication defined as
$$\forall \alpha \in \mathbb{R}, f \in C[a, b], (\alpha f)(t) = \alpha f(t) \forall t \in [a, b]$$
is a vector space.

Counter-example of vector spaces

1. Define $V = [-1, 1] \subset \mathbb{R}$. We can easily say that V is not a vector space by considering a counter-example: $1 + 1 = 2 \notin V$.

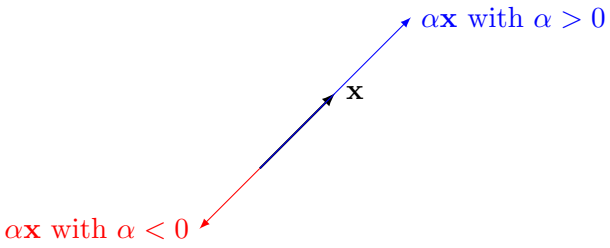
2. Consider the set of all strings with addition defined as `'I' + ' am' = 'I am'`. But this addition definition violates the commutativity of addition `"I" + " am"` $\neq$ `" am" + "I"`.
Therefore, we cannot use vector space to model text data in this naive way.

1.2 Metric on vector space

Metric on vector space is to define the "closeness/ distance" of two vectors in order to do calculus on vector spaces.
Let V be a vector space and $\mathbf{x}, \mathbf{x} \in V$, then
$$\text{dist}(\mathbf{x}, \mathbf{y}) = \text{dist}(\mathbf{x} - \mathbf{y}) = \text{length of } \mathbf{x} - \mathbf{y}$$
and (triangular inequality)
$$\text{dist}(\mathbf{x}, \mathbf{y}) + \text{dist}(\mathbf{y}, \mathbf{z}) \geq \text{dist}(\mathbf{x}, \mathbf{z})$$
Remark: distance should be *shift invariance*.
Therefore, to define distance, we only need to define the length (norm) of vectors.
Let $\mathbf{x} \in V$. Denote $||\mathbf{x}||$ be the length (norm) of $\mathbf{x}$. Then $||\mathbf{x}||$ should satisfy:

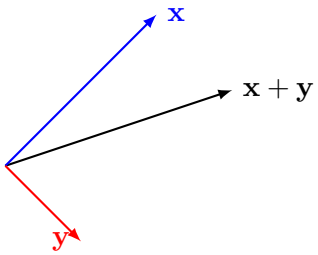
1. Non-negativity
$$||\mathbf{x}|| \geq 0$$
and
$$||\mathbf{x}|| = 0 \iff \mathbf{x} = \mathbf{0}$$

2. Length of a scaling of a vector is a scaling of the length of the vector
$$||\alpha\mathbf{x}|| = |\alpha|||\mathbf{x}||$$



3. (Triangular Inequality)

$$||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$$



Examples of vector norms

1. Euclidean norm ($2-$ norm)

$$||\mathbf{x}||_2 = (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}}$$

(a)

$$\forall i \in \{1,2,\cdots,n\}, x_i^2 \geq 0 \implies \sum_{i=1}^{n} x_i^2 \geq 0 \implies ||\mathbf{x}||_2 \geq 0$$

and

$$||\mathbf{x}||_2 = 0 \iff (\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}} = 0 \iff \sum_{i=1}^{n} x_i^2 = 0 \iff x_i = 0, \forall i \in 1,2,\cdots,n \iff \mathbf{x} = \mathbf{0}$$

(b)

$$||\alpha\mathbf{x}||_2 = (\sum_{i=1}^{n}(\alpha x_i)^2)^{\frac{1}{2}} = (\sum_{i=1}^{n}\alpha^2 x_i^2)^{\frac{1}{2}} = (\alpha^2\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}} = |\alpha|(\sum_{i=1}^{n} x_i^2)^{\frac{1}{2}} = |\alpha|||\mathbf{x}||_2$$

(c)

$$||\mathbf{x}+\mathbf{y}||_2 \leq ||\mathbf{x}||_2 + ||\mathbf{y}||_2$$

(To be proved later)

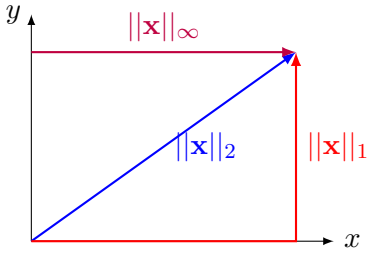
2. $\infty-$norm

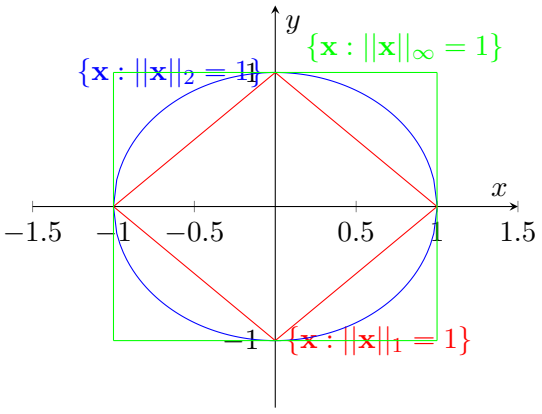
$$||\mathbf{x}||_\infty = \max_{i\in\{1,2,\cdots,n\}} |x_i|$$

3. $p-$norm

$$||\mathbf{x}||_p = (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}}$$

Comparison of $p-$norms across different p





Remark: $||\mathbf{x}||_p \leq ||\mathbf{x}||_q$ if $p \geq q$

1.3 Matrix Vector Norm

$\mathbb{R}^{m\times n}$ is a vector space. When we are talking about **matrix vector norm**, we are viewing $\mathbb{R}^{m\times n}$ as $\mathbb{R}^{mn}$ by vectorization.

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} \to \begin{bmatrix} x_{11} \\ \vdots \\ x_{m1} \\ x_{12} \\ \vdots \\ x_{m2} \\ \vdots \\ x_{1n} \\ \vdots \\ x_{mn} \end{bmatrix}$$

Then, we can define vector $p-$norm for $\mathbb{R}^{m\times n}$ as

$$||A||_{p,vec} = (\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^p)^{\frac{1}{p}}$$

For example:

1. for $p = 1$,
$$||A||_{1,vec} = (\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|)$$
2. for $p = \infty$,
$$||A||_{\infty,vec} = \max_{i,j}|a_{ij}|$$

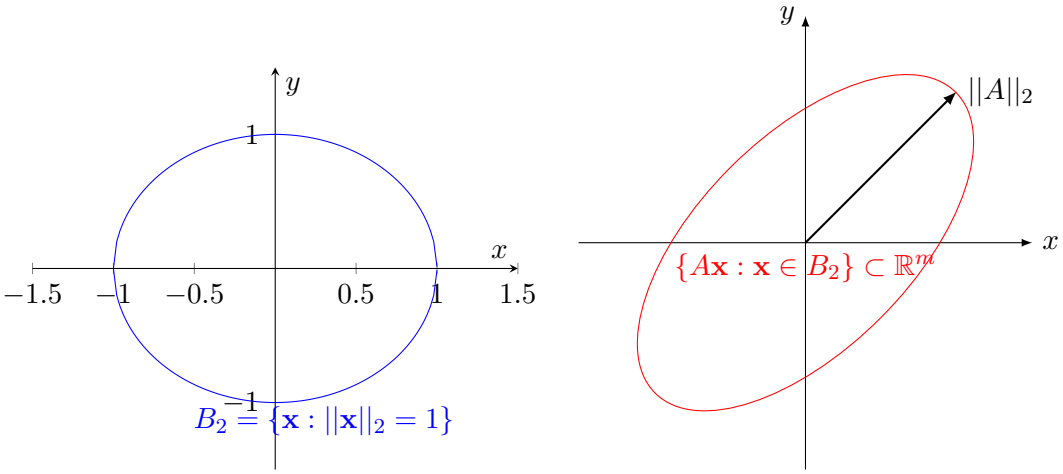
1.4 Matrix Norm

When we are talking about **matrix norm**, we are viewing $\mathbb{R}^{m\times n}$ as linear transformation $\mathbb{R}^n \mapsto \mathbb{R}^m$. Then, we define matrix $p-$norm as

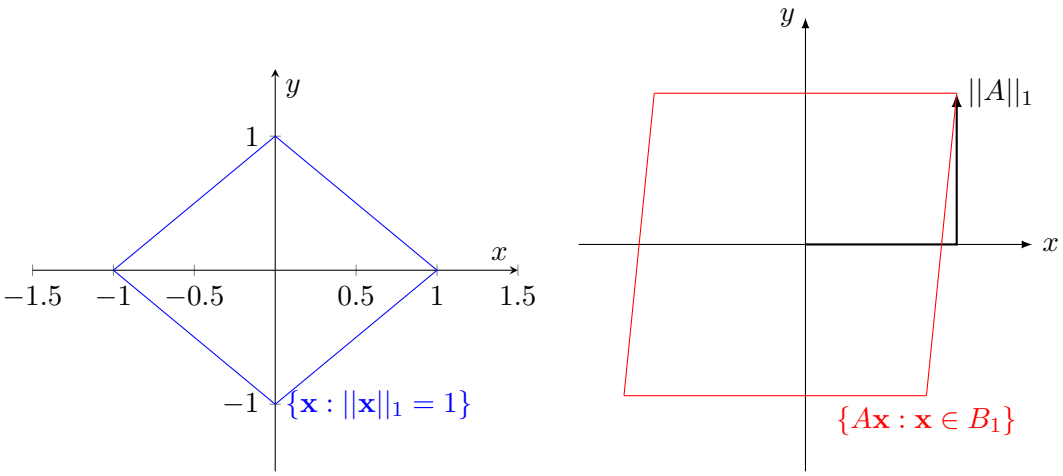
$$||A||_p = \max_{\mathbf{x}\neq 0, \mathbf{x}\in\mathbb{R}^n} \frac{||A\mathbf{x}||_p}{||\mathbf{x}||_p}$$

For example:

1. For $p = 2$,



2. For $p = 1$,



Proposition: $\|A|_-\infty = \max_{1\le i\le m} \sum_{j=1}^{\infty} |a_{ij}|$
for example

$$A = \begin{bmatrix} 2 & 1 & -1 \\ 0 & 2 & 4 \end{bmatrix} \in \mathbb{R}^{3\times 2}$$

then $\|A\|_\infty = \max |2| + |1| + |-1|, |0| + |2| + |4| = 6$

3. For $p = \infty$,

$$||A||_\infty = \max_{j\in\{1,2,\cdots,m\}} \sum_{j=1}^{n} |a_{ij}|$$

proof: it suffices to proof that

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

. We will proof that with two inequalities.

(a) $A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ then $A\mathbf{x} = \begin{bmatrix} a_1\mathbf{x} \\ a_2\mathbf{x} \\ \vdots \\ a_m\mathbf{x} \end{bmatrix}$

$$\begin{aligned} \|A\|_\infty &= \max_{1\le i\le m} |a_i x| \\ &= \max_{1\le i\le m} |\sum_{j=1}^{n} a_{ij}x_j| \\ &\le \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}||x_j| \\ &\le \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}| \end{aligned}$$

(b) Choose a special $\mathbf{x}$. Assume $\max_{1\le i\le m}\sum_{j=1}^{n}|a_{ij}| = \sum_{j=1}^{n}|a_{i'j}|$ Let $\mathbf{x} = \begin{bmatrix} sign(a_{i'1}) \\ sign(a_{i'2}) \\ \vdots \\ sign(a_{i'n}) \end{bmatrix}$

$$\begin{aligned}
\|A\mathbf{x}\|_\infty &= \max_{1\le i\le m} |\sum_{j=1}^{n} a_{ij}x_j| \\
&\ge |\sum_{j=1}^{n} a_{i'j}x_j| \\
&= |\sum_{j=1}^{n} |a_{ij}|| \\
&= \sum_{j=1}^{n} |a_{ij}| \\
&= \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|
\end{aligned}$$

By (a) and (b), we can conclude that

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

.

4. (Operator norm) For $p = 2$,

$$\begin{aligned}
||A||_2 &= \max_{||x||_2=1} ||A\mathbf{x}||_2 \\
&= \max_{||x||_2=1} ((A\mathbf{x})^T A\mathbf{x})^{\frac{1}{2}} \\
&= \max_{||x||_2=1} (\mathbf{x}^T A^T A\mathbf{x})^{\frac{1}{2}} \\
&= \text{Maximum Eigenvalues of } A^T A
\end{aligned}$$

We may also use different norms in $\mathbb{R}^n$ ($p-$norm) and $\mathbb{R}^m$ ($q-$norm).

$$||A||_{p\to q} = \max_{\mathbf{x}\neq\mathbf{0},\mathbf{x}=\mathbb{R}^n} \frac{||A\mathbf{x}||_q}{||\mathbf{x}||_p} = \max_{||\mathbf{x}||_p=1} ||A\mathbf{x}||_q$$

1.5 Other norms

1. Nuclear norm
   we can use different norms in $\mathbb{R}^n$ and $\mathbb{R}^m$.

$$||A||_{p\to q} = \max_{||\mathbf{x}||_p=1} ||A\mathbf{x}||_q$$

2. Norms on $C[a,b] = \{f : f \text{ is a continous function on } [a,b]\}$
   For all $f \in C[a,b]$, we define
   (a) $||f||_\infty = \max_{t\in[a,b]} |f(t)|$
   (b) $||f||_p = (\int_a^b |f(t)|^p dt)^{\frac{1}{p}}$ for $p \in \{1, 2, \cdots\}$
3. Norms on $l_\infty = \{\mathbf{a} : \mathbf{a} \text{ is an infinite sequence and } \exists c > 0 \text{ s.t. } |a_i| \le c\ \forall i\}$

(a) $||\mathbf{a}||_\infty = \sup_i |a_i|$

(b) $||\mathbf{a}||_p = (\sum_{i=1}^{\infty} |a_i|^p)^{\frac{1}{p}}$ but $||a||_p$ is indeed not a norm on l_∞. We will proof that by a counter-example.

$$\mathbf{a} = \begin{bmatrix} 1 \\ 1/2 \\ 1/3 \\ \vdots \\ 1/i \\ \vdots \end{bmatrix} \in l_\infty \implies ||\mathbf{a}||_1 = \sum_{i=1}^{\infty} |a_i| = \sum_{i=1}^{\infty} \frac{1}{i} = +\infty$$

If we just consider $l_p = \{\mathbf{a} : ||\mathbf{a}||_p = (\sum_{i=1}^{\infty} |a_i|^p)^{\frac{1}{p}}\}$, then we can just say $||\mathbf{a}||_p$ is a norm on l_p.
Remark: Suppose $1 \leq p < q < \infty$, then $l_p \subset l_q$

2 Case Study: K-means clustering, K-medians clustering

2.1 Clustering

Suppose we are given N vectors in $\mathbb{R}^n$,

$$x_1, x_2, \cdots, x_N \in \mathbb{R}^n$$

we are going to divide them into K different groups.
Remark: $\mathbb{R}^n$ here is used for simplicity, it can be replaced by any vector space.
Applications: image clustering, text data clustering, recommendation system, etc.

2.2 K-means clustering

Recall that `Machine Learning = Representation + Evaluation + Optimization`, in the clustering case:

1. Representation:

   (a) c_i – the group that $\mathbf{x}_i$ belongs to, for $i = 1, 2, \cdots, N$

   (b) G_j – the indices of $\mathbf{x}$'s that belongs to Group j

   $$G_j = \{i | c_i = j\}, j = 1, 2, \cdot, K$$

   $$c_i, i = 1, \cdots, N \iff G_j, j = 1, \cdots, K$$

   (c) $\mathbf{z}_j$ – the representation vector in $G_j, j = 1, \cdots K$
   Remark: $\mathbf{z}_j$ is not necessarily in $\{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$

   For example: we have $\mathbf{x}_1 = (1, 2)^T, \mathbf{x}_2 = (3, 4)^T, \mathbf{x}_3 = (4, 5)^T$, $K = 2$ and $\mathbf{x}_1, \mathbf{x}_2 \in$ Group $1, \mathbf{x}_3 \in$ Group 2. Then with the above notations:

   $$c_1 = 1, c_2 = 1, c_3 = 2$$

   $$G_1 = \{1, 2\}, G_2 = \{3\}$$

2. Evaluation:

   (a) Within Group j: we want all the vectors within the group should be close to its representation vector $\mathbf{z}_j$, that is to minimize

   $$J_j = \sum_{i \in G_j} ||\mathbf{x}_i - \mathbf{z}_j||_2^2$$

(b) All groups: we want all of J_i's to be small, that is to minimize

$$J = J_1 + J_2 + \cdots + J_K$$

Then, we want to solve the problem:

$$\min_{G_1,\cdots,G_K;\mathbf{z}_1,\cdots,\mathbf{z}_K} J = \min_{G_1,\ldots,G_K;\mathbf{z}_1,\ldots,\mathbf{z}_K} \sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2$$

This is to find $G_1,\cdots,G_K$ and $\mathbf{z}_1,\cdots\mathbf{z}_2$ that minimizes J.

3. Optimization:
In this problem, we have to find two sets of unknowns $G_1,\cdots,G_K$ and $\mathbf{z}_1,\cdots\mathbf{z}_2$. We can use alternating minimization to solve this problem.

Algorithm (Alternating minimization):
step 0: Initialize $\mathbf{z}_1,\cdots,\mathbf{z}_k$
step 1: Fix $\mathbf{z}_1,\cdots,\mathbf{z}_k$, solve the minimization with respect to $G_1,\cdots,G_K$. That is to solve

$$\min_{G_1,\ldots,G_K} \sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2 - - - - - (1)$$

step 2: Fix $G_1,\cdots,G_K$, solve the minimization with respect to $\mathbf{z}_1,\cdots,\mathbf{z}_k$. That is to solve

$$\min_{\mathbf{z}_1,\cdots,\mathbf{z}_k} \sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2 - - - - - - (2)$$

(Repeat until stopping criterion meet)

How do we solve (1), (2)?

(1)

$$\sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2 = \sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_{c_i}\|_2^2 = \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{z}_{c_i}\|_2^2$$

Therefore,

$$\begin{aligned}
\min_{G_1,\ldots,G_K} \sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2 &\iff \min_{c_1,\cdots,c_N} \sum_{i=1}^{N} \|\mathbf{x}_i - \mathbf{z}_{c_i}\|_2^2 \\
&\iff \min_{c_1,\cdots,c_N} \|\mathbf{x}_1 - \mathbf{z}_{c_1}\|_2^2 + \cdots + \|\mathbf{x}_N - \mathbf{z}_{c_N}\|_2^2 \\
&\iff \min_{c_i\in\{1,2,\cdots,K\}} \|\mathbf{x}_i - \mathbf{z}_{c_i}\|_2^2 \text{ for } i = 1,2,\cdots,N \\
&\iff c_i = \arg\min_{j\in 1,2,\cdots,K} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2
\end{aligned}$$

$\mathbf{x}_i$ is assigned to the group whose representative is the closest to $\mathbf{x}_i$. After that, we assign

$$G_j = \{i | c_i = j\} \text{ for } j = 1,\cdots,K$$

(2)

$$\sum_{j=1}^{K} \sum_{i\in G_j} \|\mathbf{x}_i - \mathbf{z}_j\|_2^2 = \sum_{i\in G_1} \|\mathbf{x}_i - \mathbf{z}_1\|_2^2 + \cdots \sum_{i\in G_K} \|\mathbf{x}_i - \mathbf{z}_K\|_2^2$$

where each term only depends on $\mathbf{z}_i$ and independent of $\mathbf{z}_1, \cdots, \mathbf{z}_{i-1}, \mathbf{z}_{i+1}, \cdots, \mathbf{z}_K$. Therefore,

$$\sum_{j=1}^{K}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_2^2 \iff \min_{\mathbf{z}_j}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_2^2 \text{ for } j=1,\cdots,K$$
$$\iff \mathbf{z}_j=\arg\min_{\mathbf{z}_j}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_2^2 \text{ for } j=1,\cdots,K$$

Consider the case in $\mathbb{R}^n$, let $f(\mathbf{z})=\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}\|_2^2$

$$\begin{aligned} f(\mathbf{z}) &= \sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}\|_2^2 \\ &= \sum_{i\in G_j}(\mathbf{x}_i-\mathbf{z})^T(\mathbf{x}_i-\mathbf{z}) \\ &= \sum_{i\in G_j}\mathbf{x}_i^T\mathbf{x}_i-2\mathbf{x}_i^T\mathbf{z}+\mathbf{z}^T\mathbf{z} \\ &= |G_j|\mathbf{z}^T\mathbf{z}+\sum_{i\in G_j}\mathbf{x}_i^T\mathbf{x}_i-2\sum_{i\in G_j}\mathbf{x}_i^T\mathbf{z} \end{aligned}$$

Then, $\nabla f(\mathbf{z})=2|G_j|\mathbf{z}-2\sum_{i\in G_j}\mathbf{x}_i$. $\nabla^2 f(\mathbf{z})=2|G_j|I$. It is obvious that $\nabla^2 f(\mathbf{z})$ is positive definite. So, $f(\mathbf{z})$ is a convex function. By setting $\nabla f(\mathbf{z})=0$, we can find the minimizer $\mathbf{z}_{minimizer}=\frac{1}{|G_j|}\sum_{i\in G_j}\mathbf{x}_i$ As a result, we can update $z_j=\frac{1}{|G_j|}\sum_{i\in G_j}\mathbf{x}_i$, which is just the mean of $\mathbf{x}_i$'s.

2.3 K-medians Clustering

In K-means Clustering, we used $2-$norm, what if we choose to use $1-$norm instead?

$$\min_{G_1,\ldots,G_K;\mathbf{z}_1,\ldots,\mathbf{z}_K}\sum_{j=1}^{K}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_1$$

This will become the *K-medians Clustering Algorithm.*

K-medians Clustering Algorithms:

step 0: Initialize $\mathbf{z}_1, \cdots, \mathbf{z}_k$

step 1: Fix $\mathbf{z}_1, \cdots, \mathbf{z}_k$, solve the minimization with respect to $G_1, \cdots, G_K$. That is to solve

$$\min_{G_1,\ldots,G_K}\sum_{j=1}^{K}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_1 \iff c_i=\arg\min_{j\in\{1,\cdots K\}}\|\mathbf{x}_i-\mathbf{z}_j\|_1 \text{ For } i=1,\cdots,N$$

step 2: Fix $G_1, \cdots, G_K$, solve the minimization with respect to $\mathbf{z}_1, \cdots, \mathbf{z}_k$. That is to solve

$$\min_{\mathbf{z}_1,\cdots,\mathbf{z}_k}\sum_{j=1}^{K}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_1 \iff \min_{j}\sum_{i\in G_j}\|\mathbf{x}_i-\mathbf{z}_j\|_1 \text{ For } j=1,\cdots,K$$

$$\mathbf{z}_j=\text{entrywise-median}\{\mathbf{x}_i|i\in G_j\} \text{ For } j=1,\cdots,K$$

(Repeat until stopping criterion meet)

2.4 Comparison of K-means and K-medians

Both mean and median can be using in clustering problem but median seems to have a better representation.

1. Mean is sensitive to outliers.
2. Median is more robust to outliers.
3. In machine learning algorithms, $1-$norm distance is more robust than $2-$norm.

3 Limit and Convergence on vector space

In calculus, we know that $\lim_{n\to\infty} a_n = a \iff \lim_{n\to\infty} |a_n - a| = 0$. Is it also true on a vector space with a norm $\|\cdot\|$? The answer is yes.

3.1 Limit and convergence on a normed vector space