

第 33 章

行列ノルム



さまざまな距離

データ量の削減などを目的に行列の近似を行ったり、誤差のある測定値を成分とする行列を扱う際には、行列に関しても誤差を考える必要が生じる。

つまり、「もとの行列にどれくらい近いか？」などという、**距離**を考える必要がある。

距離はいわば長さを測る物差しであり、一定の基準（公理）を満たせば、さまざまな距離を定義することができる。そして、使い道に応じて最適な距離を選ぶことができる。

行列の距離を考える前に、一般的な距離について考えてみよう。

距離の公理

まず、距離を名乗るものはどれも次の性質を満たすように作る必要がある。

- i. 2 点間の距離は負の数にはならない
- ii. 2 点が同じ点 \iff 2 点間の距離は 0
- iii. 2 点間の距離は行きと帰りで変わらない
- iv. 寄り道した方が必ず総距離は長い

直観的に書いた上の 4 つの性質を数学的にまとめると、次のような公理になる。

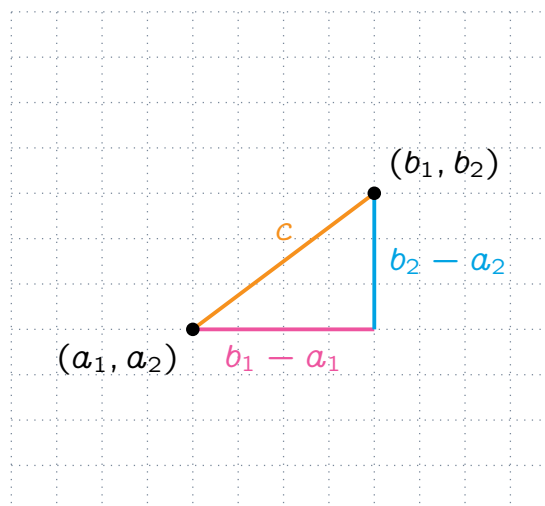
🎓 def - 距離の公理

集合 V 上の次の性質を満たす関数 $d: V \times V \rightarrow \mathbb{R}$ を距離と定める。ここで、 $a, b, c \in V$ とする。

- i. $d(a, b) \geq 0$ (非負性)
- ii. $d(a, b) = 0 \iff a = b$ (同一性)
- iii. $d(a, b) = d(b, a)$ (対称性)
- iv. $d(a, c) \leq d(a, b) + d(b, c)$ (三角不等式)

例：ユークリッド距離とマンハッタン距離

状況に応じて最適な距離が異なることを、簡単な例で考えてみよう。



次の 2 パターンの条件下において、2 点 $A = (a_1, a_2)$ と $B = (b_1, b_2)$ の距離をどう定めるべきかを考える。

1. 斜めに移動できる
2. ジグザグにしか移動できない

斜めに移動できる場合は、単純に直線距離として考えることができるので、三平方の定理を用いて、

$$d(A, B) = c = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$$

のように計算すればよい。

このような距離はユークリッド距離と呼ばれる。

一方、ジグザグにしか移動できない場合は、実際に A から B へ移動するときの遠さを直線距離では表現できないので、次のように考えた方がよい。

$$d(A, B) = |b_1 - a_1| + |b_2 - a_2|$$

このような距離はマンハッタン距離と呼ばれる。



ノルムと距離の関係

def 14.2 「ノルムの公理 (\mathbb{R} 上の線形空間)」で定めたように、ノルムは「長さ」を拡張した概念であり、ベクトルに対して定義された。

今、2 つのベクトル \mathbf{a}, \mathbf{b} を

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

とおくと、先ほど示したユークリッド距離の式

$$d(A, B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$$

は内積を経由して次のように書き換えられる。

$$d(A, B) = \sqrt{(\mathbf{b} - \mathbf{a}, \mathbf{b} - \mathbf{a})} = \|\mathbf{b} - \mathbf{a}\|$$

このように、距離からノルムを定めることができ、逆にノルムから距離を定めることもできる。

上述の式は 2 次元平面上の点の間の距離を表しているが、ベクトルのノルムとしてユークリッド距離を考えれば、 n 次元空間への拡張も容易になる。

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

とおいたとしても、

$$d(A, B) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} = \sqrt{(\mathbf{b} - \mathbf{a}, \mathbf{b} - \mathbf{a})} = \|\mathbf{b} - \mathbf{a}\|$$

が成り立つ。

L_p ノルム

ユークリッド距離とマンハッタン距離は、次の L_p ノルムによって一般化される。

🎓 def - L_p ノルム

ベクトル $\mathbf{a} = (a_i)_{i=1}^n$ に対し、

$$\|\mathbf{a}\|_p = \left(\sum_{i=1}^n |a_i|^p \right)^{\frac{1}{p}}$$

は def 14.2「ノルムの公理 (\mathbb{R} 上の線形空間)」を満たし、これを L_p ノルムという。

L_p ノルムが距離として意味を持つのは、 $p \geq 1$ のときである。

$p = 1$ の場合、

$$\|\mathbf{a}\|_1 = \sum_{i=1}^n |a_i|$$

となり、 \mathbf{a} を $\mathbf{b} - \mathbf{a}$ に置き換えれば、マンハッタン距離の式となる。

$p = 2$ の場合、

$$\|\mathbf{a}\|_2 = \left(\sum_{i=1}^n |a_i|^2 \right)^{\frac{1}{2}} = \sqrt{\sum_{i=1}^n a_i^2}$$

となり、 \mathbf{a} を $\mathbf{b} - \mathbf{a}$ に置き換えれば、ユークリッド距離の式となる。

行列の距離：フロベニウスノルム

[Todo 1:]

行列の距離：作用素ノルム

[Todo 2:]



特異値分解とベクトルのノルム

特異値分解はユニタリ変換であり、ユニタリ変換はベクトルのノルムを変えないことから、さまざまな有用な性質が導かれる。

theorem 33.1 - ユニタリ行列による座標変換

$\mathcal{U} = \{\mathbf{u}_i\}_{i=1}^n$ を \mathbb{C}^n の正規直交基底とし、ユニタリ行列 $U = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ を定める。このとき、 $\mathbf{v} \in \mathbb{C}^n$ の \mathcal{U} に関する座標ベクトルは $U^* \mathbf{v}$ で与えられる。

証明

次式が成り立つことから、 U は \mathbb{C}^n の標準基底 $\{\mathbf{e}_i\}_{i=1}^n$ から \mathcal{U} への変換行列とみなせる。

$$(\mathbf{u}_1 \ \cdots \ \mathbf{u}_n) = (\mathbf{e}_1 \ \cdots \ \mathbf{e}_n) U$$

theorem 13.1 「座標ベクトルの変換則」 より、 \mathbf{v} の \mathcal{U} に関する座標ベクトルを $\mathbf{c} \in \mathbb{C}^n$ とすると、

$$\mathbf{v} = U \mathbf{c}$$

両辺に左から U^* をかけると、

$$U^* \mathbf{v} = U^* U \mathbf{c} = \mathbf{c}$$

となり、たしかに \mathbf{c} は $U^* \mathbf{v}$ で与えられることがわかる。 

A が特異値分解されていると、ベクトルのノルムの変化がわかりやすい。

📌 theorem 33.2 - 特異値分解に基づくノルムの展開表示

行列 A がユニタリ行列 U, V を用いて $A = U\Sigma V^*$ と特異値分解されているとする。

V に対応する \mathbb{C}^n の基底を \mathcal{V} とし、 $\mathbf{v} \in \mathbb{C}^n$ の \mathcal{V} に関する座標ベクトルを $\mathbf{c} = (c_i) \in \mathbb{C}^n$ とすると、

$$\|A\mathbf{v}\|^2 = \sum_{i=1}^r \sigma_i^2 |c_i|^2$$

🔪 証明

theorem 33.1「ユニタリ行列による座標変換」より、座標ベクトルは $\mathbf{c} = V^*\mathbf{v}$ で与えられる。

よって、

$$\|A\mathbf{v}\|^2 = \|U\Sigma V^*\mathbf{v}\|^2 = \|U\Sigma\mathbf{c}\|^2$$

ここで、theorem 20.6「ユニタリ行列の特徴づけとしてのノルム不変性」より、左からユニタリ行列 U をかけてもノルムは変わらないので、

$$\|A\mathbf{v}\|^2 = \|\Sigma\mathbf{c}\|^2$$

Σ は対角成分に特異値 $\sigma_1, \dots, \sigma_r$ ($\sigma_i > 0$) を持ち、それ以外は 0 の $m \times n$ 行列であるから、

$$\Sigma\mathbf{c} = (\sigma_1 c_1, \dots, \sigma_r c_r, 0, \dots, 0)^T$$

よってそのノルムの二乗は、def 14.4「 \mathbb{C}^n 上の内積（標準内積）」より、 \mathbb{C}^n 上の自分自身との内積と考えて、

$$\|\Sigma\mathbf{c}\|^2 = \sum_{i=1}^r |\sigma_i c_i|^2 = \sum_{i=1}^r \sigma_i^2 |c_i|^2$$

となり、目的の式が示された。 ■



最大特異値と作用素ノルム

A の最大特異値 σ_1 は、次の解釈をもつ。

theorem - 最大特異値と作用素ノルムの一致

複素行列 A に対して、

$$\max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\| = \sigma_1$$

証明

theorem 33.2「特異値分解に基づくノルムの展開表示」と、 $\sigma_i \leq \sigma_1$ を用いて、

$$\|A\mathbf{v}\|^2 = \sum_{i=1}^r \sigma_i^2 |c_i|^2 \leq \sum_{i=1}^r \sigma_1^2 |c_i|^2 = \sigma_1^2 \sum_{i=1}^r |c_i|^2$$

ここで、 $\|\mathbf{v}\| = 1$ ならば、 $\mathbf{v} \in \mathbb{C}^n$ の \mathcal{V} に関する座標ベクトル \mathbf{c} も $\|\mathbf{c}\| = 1$ となるので、

$$\|A\mathbf{v}\|^2 \leq \sigma_1^2 \sum_{i=1}^r |c_i|^2 = \sigma_1^2 \|\mathbf{c}\|^2 = \sigma_1^2$$

よって、 $\|A\mathbf{v}\| \leq \sigma_1$ が任意の単位ベクトル \mathbf{v} に対して成り立つ。

すなわち、

$$\max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\| \leq \sigma_1$$

等号は、 $c_1 = 1$ で他の成分 c_2, \dots, c_r が 0 のときに成り立つ。 ■

def - 作用素ノルム

複素行列 A に対して、 A の作用素ノルムを次のように定義する。

$$\|A\| := \max_{\|\mathbf{v}\|=1} \|A\mathbf{v}\|$$

.....

Zebra Notes

Type	Number
todo	2