第 VII 部　市場の分析

7　$\boldsymbol{n}$ 次元空間の距離

ポイント

- ベクトル
- ノルム
- n 次元空間の距離
- K-means

7.1　はじめに

- 距離の概念を n 次元空間で使えるように定義しなおします。
- 定義は受け入れるもので理解するものです。
- 何を言ってるのかを理解してください。

7.2　ベクトル

定義

n 個の実数 $a_1, a_2, \cdots, a_n$ を、この順序で並べた

$$\boldsymbol{a} = (a_1, a_2, \cdots, a_n) \tag{1}$$

をベクトルとよび、太字の小文字 ($\boldsymbol{a}$) を使ってあらわす。このとき、a_i を『成分』とよぶ。

$a_1, a_2, \cdots, a_n$ をそれぞれ第 1 成分、第 2 成分、$\cdots$ 第 n 成分とよぶ。成分の個数が n 個であるベクトルを『n 次元ベクトル』とよぶ

定義

同じ次元の 2 つのベクトル

$$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$$

$$\boldsymbol{b} = (b_1, b_2, \cdots, b_n)$$

において対応する全ての成分が等しいときに限って『等しい』と定義し、

$$\boldsymbol{a} = \boldsymbol{b} \tag{2}$$

とあらわす。

定義

同じ次元の 2 つのベクトル

$$\boldsymbol{a} = (a_1, a_2, \cdots, a_n), \qquad \boldsymbol{b} = (b_1, b_2, \cdots, b_n)$$

の和を $\boldsymbol{a}+\boldsymbol{b}$ であらわし

$$\boldsymbol{a}+\boldsymbol{b} = (a_1+b_1\ ,\ a_2+b_2\ ,\ \cdots\ ,\ a_n+b_n) \tag{3}$$

と定義する。

定義

ベクトル $\boldsymbol{a}$ にスカラー α をかけるということを $\alpha\boldsymbol{a}$ とあらわし

$$\alpha\boldsymbol{a} = \alpha(a_1\ ,\ a_2\ ,\ \cdots\ ,\ a_n) \tag{4}$$

$$= (\alpha a_1\ ,\ \alpha a_2\ ,\ \cdots\ ,\ \alpha a_n) \tag{5}$$

と定義する。

定義

全ての成分が 0 であるベクトル

$$\mathbf{0} = (0, 0, \ldots, 0) \tag{6}$$

をゼロ・ベクトルとよび、$\mathbf{0}$ であらわす。

問題 VII–7 – 1

　同じ次元の 2 つのベクトルを

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$$

$$\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$$

とする

$$\boldsymbol{a} - \boldsymbol{b}$$

を求めなさい

解例 VII–7 – 1

$$
\begin{aligned}
\boldsymbol{a}-\boldsymbol{b} &= (a_1, a_2, \ldots, a_n) - (b_1, b_2, \ldots, b_n) \\
&= (a_1, a_2, \ldots, a_n) + \big((-1)(b_1, b_2, \ldots, b_n)\big) \\
&= (a_1, a_2, \ldots, a_n) + (-b_1, -b_2, \ldots, -b_n) \\
&= (a_1 - b_1, a_2 - b_2, \cdots, a_n - b_n)
\end{aligned}
$$

7.2.1　ベクトルの内積

定義

2 つの n 次元ベクトル

$$\boldsymbol{a} = (a_1\ ,\ a_2\ ,\ \ldots\ ,\ a_n)$$

$$\boldsymbol{b} = (b_1\ ,\ b_2\ ,\ \ldots\ ,\ b_n)$$

に対して、

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \tag{7}$$

によって定まる数値 $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ を、$\boldsymbol{a}$ と $\boldsymbol{b}$ の『内積』という。

定理 VII–7 – 1

内積は以下の性質を持つ。

1. $\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \langle \boldsymbol{b}, \boldsymbol{a} \rangle$
2. $\langle \boldsymbol{a}, \boldsymbol{a} \rangle \geqq 0$

証明 VII–7 – 1.1

$\boldsymbol{a} = (a_1, a_2, \cdots, a_n), \boldsymbol{b} = (b_1, b_2, \cdots, b_n)$ とする。内積の定義により

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = a_1b_1 + a_2b_2 + \cdots + a_nb_n \tag{8}$$

$$\langle \boldsymbol{b}, \boldsymbol{a} \rangle = b_1a_1 + b_2a_2 + \cdots + b_na_n \tag{9}$$

であって、右辺各項は交換法則により

$$a_ib_i = b_ia_i \tag{10}$$

が成り立つから、その和である右辺は等しい。したがって、

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle = \langle \boldsymbol{b}, \boldsymbol{a} \rangle \tag{11}$$

が成り立つ。（終）

証明 VII–7 – 1.2

$\boldsymbol{a} = (a_1, a_2, \cdots, a_n)$ とする。内積の定義により

$$\langle \boldsymbol{a}, \boldsymbol{a} \rangle = a_1 a_1 + a_2 a_2 + \cdots + a_n a_n = {a_1}^2 + {a_2}^2 + \cdots + {a_n}^2 \tag{12}$$

であり、最右辺各項は二乗なので、最右辺全ての項に対して

$${a_i}^2 \geqq 0 \tag{13}$$

が成り立つから、その和は非負である。したがって、

$$\langle \boldsymbol{a}, \boldsymbol{a} \rangle \geqq 0 \tag{14}$$

が成り立つ。（終）

7.2.2　ノルム

定義

$\sqrt{\langle \boldsymbol{a},\ \boldsymbol{a} \rangle}$ を $\boldsymbol{a}$ の『ノルム』といい、$\|\boldsymbol{a}\|$ であらわす。

$\boldsymbol{a} = (a_1,\ a_2,\ \cdots,\ a_n)$ のとき、

$$\|\boldsymbol{a}\| = \sqrt{\langle \boldsymbol{a},\ \boldsymbol{a} \rangle} = \sqrt{{a_1}^2 + {a_2}^2 + \cdots + {a_n}^2} \tag{15}$$

である。

問題 VII–7 – 2

同じ次元の 2 つのベクトルを

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_n)$$

$$\boldsymbol{b} = (b_1, b_2, \ldots, b_n)$$

とする

$$\boldsymbol{a} - \boldsymbol{b}$$

のノルムを求めなさい

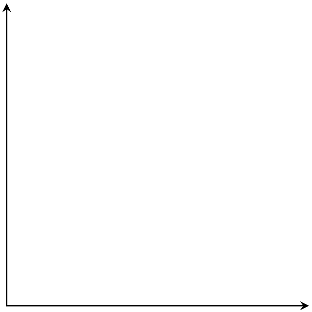
解例 VII–7 – 2

$$\langle \boldsymbol{a}-\boldsymbol{b}, \boldsymbol{a}-\boldsymbol{b}\rangle = (a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_n-b_n)^2$$
$$\|\boldsymbol{a}-\boldsymbol{b}\| = \sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_n-b_n)^2}$$

7.3　ベクトルと座標

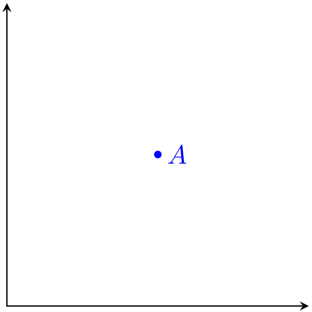
平面上に直交する 2 本の直線を描き、それらを座標軸として平面の座標とする。

図 1　直交する二本の直線



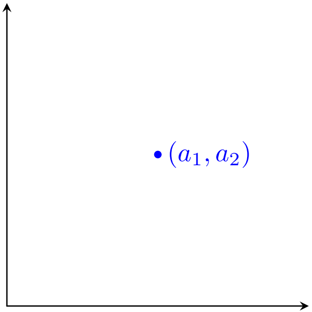
平面上に点 A をとる。

図 2　平面上の点



これにより、平面上の各点に座標 (a_1, a_2) が与えられる。

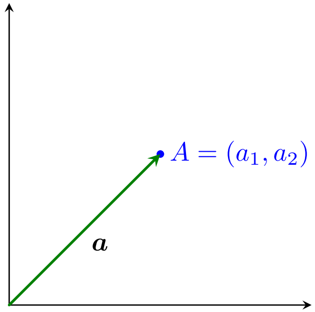
図 3　点 A の座標



この (a_1, a_2) を $\boldsymbol{a} = (a_1, a_2)$ と書けば $\boldsymbol{a}$ が座標 (a_1, a_2) をもつ平面上の点 A に対応する。

このとき、$\boldsymbol{a}$ を原点から A への有向線分であらわすと、$\boldsymbol{a}$ と平面上の点 A さらに (a_1, a_2) を対応させることができる。

図 4　座標上のベクトル



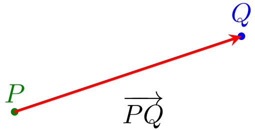
7.3.1 n 次元への拡張

- 同様に、直交する 3 本の直線を用いれば、3 次ベクトル $\boldsymbol{a} = (a_1, a_2, a_3)$ は 3 次元における座標 (a_1, a_2, a_3) に対応する。
- この考え方は n 次元へと拡張できる。

7.4 有向線分

- 平面上に任意の点 P, Q を取り線分 PQ に、P から Q に向かう方向を与える。
- 方向を与えられた線分 PQ を有向線分といい $\overrightarrow{PQ}$ であらわす。
- P を『始点』と呼び、Q を『終点』と呼ぶ。
- $\overrightarrow{PQ}$ の大きさは線分 PQ の長さで示す。

図 5　有向線分



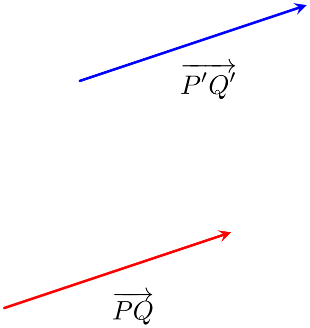
7.4.1　等しい有向線分の組

二つの有向線分 $\overrightarrow{PQ}, \overrightarrow{P'Q'}$ が同じ方向を持ち、かつ長さも等しいとき、これら二つの有向線分は『等しい』といい、このことを

$$\overrightarrow{PQ} = \overrightarrow{P'Q'} \tag{16}$$

であらわす。

図 6　等しい有向線分



7.5　幾何ベクトル

等しい有向線分どうしを一つの組にまとめ各組を $\boldsymbol{a}, \boldsymbol{b}, \cdots$ 等の文字であらわすことにする。$\overrightarrow{PQ}$ に等しい有向線分の組が $\boldsymbol{a}$ であるとき

$$\boldsymbol{a} = \overrightarrow{PQ} \tag{17}$$

と書く。$\overrightarrow{PQ} = \overrightarrow{P'Q'}$ であれば、

$$\boldsymbol{a} = \overrightarrow{P'Q'} \tag{18}$$

であり、$\overrightarrow{PQ}$ に等しい有向線分はすべて $\boldsymbol{a}$ であらわされる。このように定義された有向線分の組 $\boldsymbol{a}, \boldsymbol{b}, \cdots$ を『幾何ベクトル』という。

7.5.1　ゼロ・ベクトル

- 始点と終点が同じ点である $\overrightarrow{PP}$ を有向線分の仲間に入れる。
- $\overrightarrow{PP}$ の長さはゼロであり、方向は持たない。
- このような $\overrightarrow{PP}$ に等しい有向線分の組を『ゼロ・ベクトル』と呼び、$\mathbf{0}$ であらわす。

$$\mathbf{0} = \overrightarrow{PP} \tag{19}$$

7.5.2　位置ベクトル

- 平面上に一点 O をとりこの点を固定する。
- 幾何ベクトル $\boldsymbol{a}$ のなかから O を始点とする有向線分

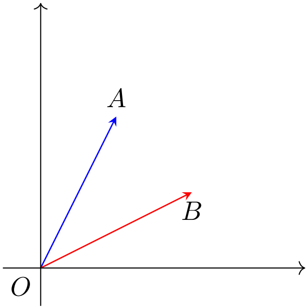
$$\boldsymbol{a} = \overrightarrow{OP} \tag{20}$$

  を選ぶ。
- この $\overrightarrow{OP}$ を始点を O に固定した時の『位置ベクトル』とよぶ。

7.5.3　幾何ベクトルの和

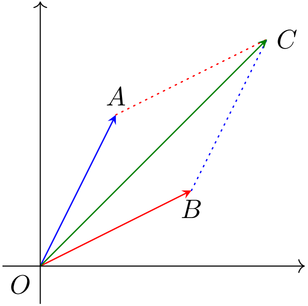
- 二つの幾何ベクトル $\boldsymbol{a}, \boldsymbol{b}$ をとる。
- 平面上の点 O をとり $\boldsymbol{a} = \overrightarrow{OA}, \boldsymbol{b} = \overrightarrow{OB}$ となる点 A と B をとる。

図 7　二つの幾何ベクトル



- 線分 OA,OB を隣り合う辺とする平行四辺形を作り、この平行四辺形のもう一つの頂点を C とする。
- 点 O を始点とし、点 C を終点とする有向線分 $\overrightarrow{OC}$ で与えられる幾何ベクトルを $\boldsymbol{c}$ とする
- この $\boldsymbol{c}$ を $\boldsymbol{a}$ と $\boldsymbol{b}$ の和と定義し $\boldsymbol{c} = \boldsymbol{a} + \boldsymbol{b}$ とあらわす。

図 8　幾何ベクトルの和



7.5.4 ゼロ・ベクトルのスカラー倍

幾何ベクトル $\boldsymbol{a}$ と実数 α の積において $\boldsymbol{a} = \mathbf{0}$ であれば全ての実数 α において

$$\alpha\boldsymbol{a} = \mathbf{0} \tag{21}$$

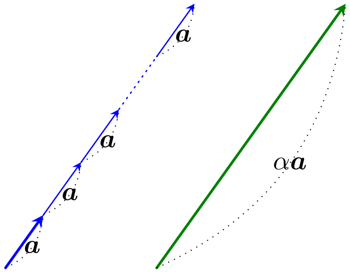
と定義する。

7.5.5　幾何ベクトルのスカラー倍

- $\boldsymbol{a} \neq \boldsymbol{0}$ とする。
- α >0 のとき

  $\boldsymbol{a}$ と同じ方向を持ち、長さが $\boldsymbol{a}$ の長さの α 倍である有向線分によってあらわされる幾何ベクトルを $\alpha\boldsymbol{a}$ とあらわす。

図 9　$\alpha > 0$



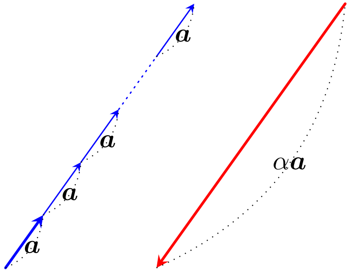
- $\alpha < 0$ のとき

  $\boldsymbol{a}$ と反対の方向を持ち、長さが $\boldsymbol{a}$ の長さの $|\alpha|$ 倍である有向線分によってあらわされる幾何ベクトルを $\alpha\boldsymbol{a}$ とあらわす。

- $\alpha = 0$ のとき

  全ての $\boldsymbol{a}$ に対して、$\alpha\boldsymbol{a} = \boldsymbol{0}$ と定義する。

図 10 $\alpha < 0$



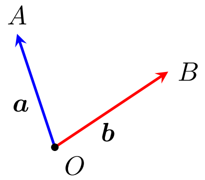
7.5.6　ベクトルの差

$\boldsymbol{a} = \overrightarrow{OA}, \boldsymbol{b} = \overrightarrow{OB}$ とする。

$$\boldsymbol{a} - \boldsymbol{b} \tag{22}$$

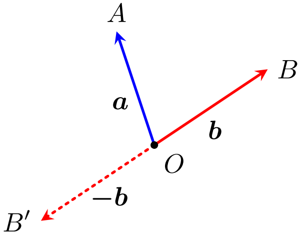
を考える。

図 11　$\boldsymbol{a}$ と $\boldsymbol{b}$



$-\boldsymbol{b}$ は $\overrightarrow{OB}$ と同じ長さを持ち向きが反対の幾何ベクトルである。O を対称の点として、B に対応する点をとり B' とする。

図 12　$-\boldsymbol{b}$



$-\boldsymbol{b} = \overrightarrow{OB'}$ なので

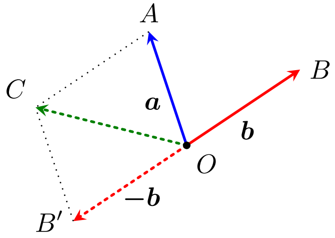
$$\boldsymbol{a} - \boldsymbol{b} = \overrightarrow{OA} - \overrightarrow{OB} \tag{23}$$

$$= \overrightarrow{OA} + \overrightarrow{OB'} \tag{24}$$

(24) はベクトルの和なので、 $\overrightarrow{OA}, \overrightarrow{OB'}$ を隣り合う辺とする平行四辺形のもう一方の頂点を C とすると

$$\overrightarrow{OA} + \overrightarrow{OB'} = \overrightarrow{OC} \tag{25}$$

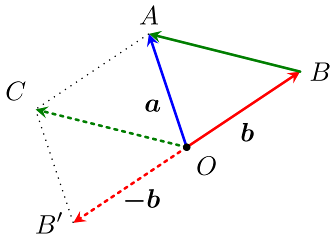
図 13　$\boldsymbol{a} + -\boldsymbol{b}$



$\boldsymbol{c} = \overrightarrow{OC}$ とすると

$$\boldsymbol{a} - \boldsymbol{b} = \boldsymbol{c} = \overrightarrow{OC} = \overrightarrow{BA} \tag{26}$$

図 14　$\boldsymbol{a} - \boldsymbol{b}$



問題 VII–7 – 3

$\boldsymbol{a} = \overrightarrow{OA}, \boldsymbol{b} = \overrightarrow{OB}$ とする。

$\boldsymbol{a} = (a_1, a_2)$,$\boldsymbol{b} = (b_1, b_2)$ のとき以下の問いに答えなさい。

1. A の座標を求めなさい。
2. B の座標を求めなさい。
3. O を対称の点として、B に対応する点をとり B' とする。B' の座標を求めなさい。
4. $\overrightarrow{OA}, \overrightarrow{OB'}$ を隣り合う辺とする平行四辺形のもう一方の頂点を C とする。C の座標を求めなさい。
5. $\boldsymbol{c} = \boldsymbol{a} - \boldsymbol{b}$ のノルムを求めなさい。

解例 VII–7 – 3

1. (a_1, a_2)
2. (b_1, b_2)
3. $(-b_1, -b_2)$
4. $(a_1 - b_1, a_2 - b_2)$
5. $\|\boldsymbol{c}\| = \|\boldsymbol{a} - \boldsymbol{b}\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$

図 15　$\boldsymbol{a}$ と $\boldsymbol{b}$

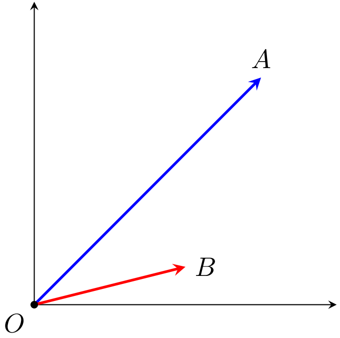


図 16　$\boldsymbol{a}-\boldsymbol{b}$

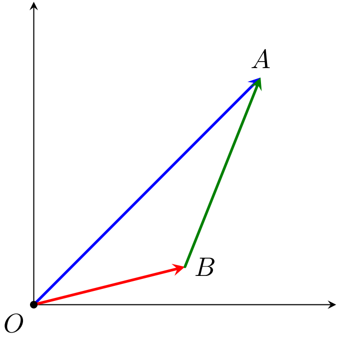


図 17　AB の座標

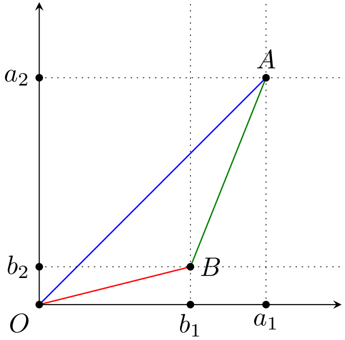
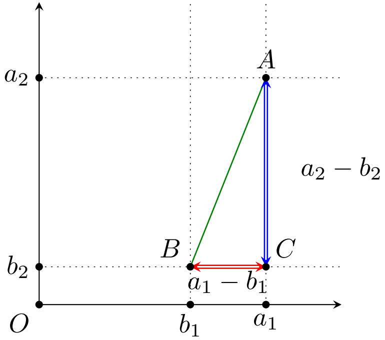


図 18　ピタゴラスの定理



問題 VII–7 – 4

$\boldsymbol{a} = \overrightarrow{OA}, \boldsymbol{b} = \overrightarrow{OB}$ とする。

$\boldsymbol{a} = (a_1, a_2, a_3)$,$\boldsymbol{b} = (b_1, b_2, b_3)$ のとき以下の問いに答えなさい。

1. A の座標を求めなさい。
2. B の座標を求めなさい。
3. O を対称の点として、B に対応する点をとり B' とする。B' の座標を求めなさい。
4. $\overrightarrow{OA}, \overrightarrow{OB'}$ を隣り合う辺とする平行四辺形のもう一方の頂点を C とする。C の座標を求めなさい。
5. $\boldsymbol{c} = \boldsymbol{a} - \boldsymbol{b}$ のノルムを求めなさい。

解例 VII–7 – 4

1. (a_1, a_2, a_3)
2. (b_1, b_2, b_3)
3. $(-b_1, -b_2, -b_3)$
4. $(a_1 - b_1, a_2 - b_2, a_3 - b_3)$
5. $\|\boldsymbol{c}\| = \|\boldsymbol{a} - \boldsymbol{b}\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}$

定義

$\boldsymbol{a} = \overrightarrow{OA}, \boldsymbol{b} = \overrightarrow{OB}$ において、$\boldsymbol{c} = \overrightarrow{BA}$ とすると、

$$\overrightarrow{BA} = \boldsymbol{c} = \boldsymbol{a} - \boldsymbol{b} \tag{27}$$

なので、点 A, B の距離を $\boldsymbol{c}$ のノルムすなわち

$$\|\boldsymbol{c}\| = \|\boldsymbol{a} - \boldsymbol{b}\| \tag{28}$$

$$= \sqrt{\langle \boldsymbol{a} - \boldsymbol{b}\ ,\ \boldsymbol{a} - \boldsymbol{b} \rangle} \tag{29}$$

$$= \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2} \tag{30}$$

で定義する。この距離をユークリッド距離という。

7.6 K-means の手順

1. 予めクラスタ数を決定する。
2. ランダムに初期値となる各クラスタの中心点を布置。
3. 各データを最も近い中心点を持つクラスタに分類。
4. 所属するデータの算術平均を中心点とする。
5. 3.4. を中心点が移動しなくなるまで繰り返す。

7.6.1 K-means の特徴

- 予めクラスタ数を決定する。
- 初期値により最終クラスタが変わることがある。

7.7 まとめ

- n 次元空間における 2 点間の距離をノルムで定義する。
- $\|\boldsymbol{a}-\boldsymbol{b}\|=\sqrt{(a_1-b_1)^2+(a_2-b_2)^2+\cdots+(a_n-b_n)^2}$
- K-means
  - 予めクラスター数を指定。
  - もっとも近いクラスター中心を基準。
  - クラスターの中心は平均で算出（重心）。
  - 初期値により結果が異なる場合がある。