

線形代数学入門

まい.

概要

本資料では、機械学習や深層学習で頻繁に活用される線形代数学の基礎となる事項を述べる。

なるべく平素な言葉を使うことを心がけ、数学の知識が少ない人にでも読みやすくなるように工夫をした (つもりです)。線形代数学とは何かというところや行列から始め、簡単な例 (小さなサイズの行列) を通して線形代数学の基礎的な知識を得られるようにした。線形代数学について真面目に書こうとすると、ページ数が膨大となり終わりが見えなくなるためなるべく定義や定理などを挿入しない形をとった (真面目に全部書くとそれこそ「本」を書くことになるので)。

この資料を読んだ人たちが、線形代数学がどんなものか雰囲気を感じ取ってもらい、ちゃんとした数学書などで勉強しようと意欲を持てば何よりも嬉しいです。

目次

1	線形代数学とは	3
1.1	線形とは	3
1.2	代数とは	3
1.3	改めて線形代数学とは	3
2	ベクトルと行列	4
3	行列の演算	5
3.1	行列の加法と減法	5
3.2	行列の乗法	5
4	行列と連立方程式	7
4.1	行列で連立方程式を記述する	7
4.2	行列を使って連立方程式を解く	8
5	行列の幾何的な意味	10
6	固有値と固有ベクトル	10
6.1	固有値と固有ベクトルとは	11
6.2	固有値と固有ベクトルの求め方	11
7	行列の対角化	13
7.1	対角行列と冪乗計算	13
7.2	行列の対角化	14
7.3	行列の対角化の応用	16
8	行列の特異値分解	17
8.1	特異値分解とは	17
8.2	特異値分解の計算方法 (概略)	17
9	おまけ：行列の Jordan 標準形	18

1 線形代数学とは

1.1 線形とは

線形とは平たい言葉で言うと”まっすぐ”のことである。小学生から定規で一度は引いたことはあるだろうあの”まっすぐ”である。小学生から学ぶということは、図形の中でおそらく最も扱いやすく簡単であるということだ。

まっすぐな図形と聞くときっと、小学生もしくは中学生で学んだ「比例」の式や「一次関数」の式である

$$y = ax + b \ (a \neq 0)$$

を思い浮かべる人がいるだろう。まさにこの式こそが線形であることを述べている。

つまり、「線形」とは「まっすぐ」を数学的に言い表したに過ぎない (が線形代数学が簡単と言っているわけではない)。

1.2 代数とは

代数とは「数学の一分野で、数の代わりに文字を用いて方程式の解法などを研究する学問」*1のことである。例えば、次のような x, y に関する方程式に関して研究する分野のことを一般には代数学という*2:

$$\begin{aligned} 2x + 1 &= 2, \\ 3x^2 - x - 2 &= 0, \\ a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 &= 0 \ (a_i \in \mathbb{R}), \\ \begin{cases} 2x + y = 2 \\ x - y = 1 \end{cases} \end{aligned}$$

1.3 改めて線形代数学とは

上記から改めて線形代数学というものを考えると、

まっすぐを中心とした理論を方程式の観点から研究する分野*3

であると言えるだろう。また、代数学において線形とは、いわゆる 1 次式のことを表しており、つまりは 1 次方程式に関する学問だということもできる。今では数学の基礎となっており、線形代数学なしで数学を語ることはできない。

それでは、まっすぐな数学の世界を楽しんでください。

*1 wikipedia 参照

*2 もちろんこれだけでなくさまざまな対象を研究している

*3 ちゃんとした定義などは本やネットで調べてみてください

2 ベクトルと行列

本節では、ベクトルと行列について述べる。「ベクトル」とは高校のときの数学 B で学んだあのベクトルのことであるが、「行列」とはお店で並ぶ列のことではない。実はベクトルも行列の一部であり、

行列はベクトルを抽象化したもの。

と言えるだろう。

ベクトルとは、「向き」と「大きさ」をもった量のことをいい、これも「まっすぐ」な数学の対象である。よく次のように成分表示される：

$$\vec{a} = (1, 2).$$

このベクトルは、数が 1 行 2 列に並んでいると思って欲しい。すると行列とは、この行と列の数をもっと増やしたものと抽象化できる。例えば、次のようなものが行列である：

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} -1 & 0 & 8 \\ 0 & 2 & 2 \\ 10 & 3 & 0 \\ 0 & 1 & 6 \end{bmatrix}, C = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}$$

行列はよく大文字のアルファベットを用いて表される。また、数学界には行列を書くときに丸括弧にするのか四角括弧にするのかの二大派閥が存在するが、筆者は四角括弧派の人間である。^{*4}

行列で大事なことはその中の数よりも、行数と列数である。例えば、上記の行列 A は 2 行 2 列の行列、 B は 4 行 3 列の行列、 C は 4 行 1 列の行列ということになる。また、行と列が等しい行列のことを特別に正方行列と呼ぶ。つまり、行列 A は正方行列であり、特に 2 次正方行列という（一般に n 次正方行列がある）。

行列のサイズを述べるときは n 行 m 列の行列であれば

$$n \times m \text{ 行列}$$

ということがある。

行列の中の数のことを「成分」と呼ぶ。行と列を指定することでその数を抽出できる。例えば、上記の行列 A の $(1, 2)$ 成分は 2 である。他にも、行列 B の $(3, 2)$ 成分は 3 であるとわかる。

^{*4} 大きな行列を手書きするときには曲線より直線で描く四角括弧の方が綺麗にかけるため

3 行列の演算

本節では、行列の計算方法について述べる。足し算・引き算は非常に直感的でなんの違和感もない計算方法だが、掛け算は慣れていないと少し難しく感じるかもしれない。

3.1 行列の加法と減法

行列の加法と減法は、成分ごとに計算をすれば良いので非常に簡単である。ただし、計算する2つの行列が全く同じサイズでなければならないことに注意が必要である。具体的には次のように計算ができる：

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix}$$

のとき

$$\begin{aligned} A + B &= \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 6 & 0 \end{bmatrix} \end{aligned}$$

のように成分ごとに計算することができる。もちろん、減法に関しても

$$\begin{aligned} A - B &= \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & -4 \\ 3 & 2 & 0 \end{bmatrix} \end{aligned}$$

と成分ごとに引き算をすればよい。

3.2 行列の乗法

行列の乗法に関しては少し計算方法がややこしい。習得してもらうためにはまずベクトルの内積を復習する。ベクトルの内積とは、2つのベクトルに対して定義される特別な演算で、

$$\begin{aligned} \vec{a} &= (x_1, y_1), \\ \vec{b} &= (x_2, y_2) \end{aligned}$$

という2つのベクトルに対して、

$$\vec{a} \cdot \vec{b} = x_1 x_2 + y_1 y_2$$

と各成分を掛けて足すという計算をした。この計算方法が行列の乗法でも使われる。

3.2.1 行列のスカラー倍

行列にスカラーをかけると、各成分がそのスカラー倍されることになる。例えば次のようにスカラー倍は計算できる:

$$3 \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ 3 & -3 \end{bmatrix}$$

いたって普通の計算である。

3.2.2 行列の普通の積

行列の一般的な積の計算は次のように行われる。細かく計算過程を記載するのでじっくりと眺めてみてほしい。行列

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 \\ -1 & 2 \\ 1 & 2 \end{bmatrix}$$

に対して、行列の積を次のように計算する:

$$\begin{aligned} AB &= \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 2 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 \cdot 0 + 0 \cdot (-1) + (-2) \cdot 1 & 1 \cdot 1 + 0 \cdot 2 + (-2) \cdot 2 \\ 2 \cdot 0 + 4 \cdot (-1) + 0 \cdot 1 & 2 \cdot 1 + 4 \cdot 2 + 0 \cdot 2 \end{bmatrix} \\ &= \begin{bmatrix} -2 & -3 \\ -4 & 10 \end{bmatrix} \end{aligned}$$

赤色で示した部分をベクトルの内積と思って計算すると1つの要素がわかる。(1,2)要素を計算したいとすれば、行列Aの1行目と行列Bの2列目をそれぞれベクトルだと思って内積を計算すれば-3が計算できる。かなり複雑で慣れていないと素早く計算ができないが、内積計算の繰り返しで成分の場所にさえ気をつければ問題なく慣れることだろう。また、行列の積は順序によって計算結果が異なることにも注意が必要である。

さらに、行列の積が計算できる場合は、左側の行列の列数と右側の行列の行数が一致している必要がある。そして行列の積のサイズは「左の行列の行数」×「右の行列の列数」となる(上の例であれば2×2となる)。

逆に、行列を2つの行列の積に書き直すことも軽く練習しておこう。次の2行1列の行列は以下のように変形することが可能である:

$$\begin{bmatrix} 2x + y \\ x - y \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

3.2.3 アダマール積

アダマール積とは、普通の数学の行列の計算で使われることはあまりないが、機械学習分野ではよく使われる計算方法である。先ほど紹介した掛け算と比較すると全く難しくな

い。成分ごとに掛け算をするという「本来こっちがメインの積だろう」と突っ込んでしまいたくなる掛け算がアダマール積である。つまり、行列

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix}$$

に対して、

$$\begin{aligned} A \odot B &= \begin{bmatrix} 1 & 0 & -2 \\ 2 & 4 & 0 \end{bmatrix} \odot \begin{bmatrix} 0 & 1 & 2 \\ -1 & 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & -4 \\ -2 & 8 & 0 \end{bmatrix} \end{aligned}$$

のように計算することをアダマール積と呼ぶ。この計算を見てわかるように、2つの行列のサイズが同じでないと計算はできない。

4 行列と連立方程式

本節では、行列と連立方程式の関係について述べる。ここからがいよいよ線形代数学になると言ってもよいだろう。行列が代数学(方程式を解く)でどのように使われていて、どれだけ強力な道具となるか体感してもらいたい。

4.1 行列で連立方程式を記述する

次の2元1次の連立方程式を考える。

$$\begin{cases} 2x + y = 2 \\ x - y = 4 \end{cases}$$

この連立方程式を見ると読者は解きたくて解きたくてたまらなくなるはずだ。この連立方程式を解くと解は

$$(x, y) = (2, -2)$$

となったはずだ。この解は特に重要ではないので、再び方程式に目を向けてもらいたい。この方程式で重要な部分とは一体どこだろうか？ x, y に関してはただの文字なので、正直これが「あ」「い」という文字だとして

$$\begin{cases} 2\text{あ} + \text{い} = 2 \\ \text{あ} - \text{い} = 4 \end{cases}$$

という方程式を考えても解は先ほどと同じになる。くどいようだが、 x, y を「afhgf」, 「Fuefr」と意味不明な文字列にしても解は導出可能だ。つまり、方程式において重要なことは x, y のような変数ではなく、2, 1などの係数が最も重要ということだ。なんなら左についている中括弧みたいなものも本質的には不要である。

方程式の中の最も重要な対象である係数のみに着目できるように、連立方程式を次のように行列で表記する:

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

また、左辺の左側の行列を A 、右側の行列を x 、右辺の行列を b とおくと次のようにさらに簡潔に書ける:

$$Ax = b$$

この式の形はまさに「線形」である。上記は 2 元の連立方程式で記述したが、これは一般に n 元の連立方程式でも同様に記述ができる。行列を用いることで連立方程式はただの 1 次方程式へと変貌する。

4.2 行列を使って連立方程式を解く

中学生が学ぶような次の方程式には、いわゆる解の公式が存在する:

$$ax = b \Leftrightarrow x = \frac{b}{a} \quad (a \neq 0)$$

では、行列で書かれた連立方程式も

$$Ax = b \Leftrightarrow x = \frac{b}{A}$$

のようにシステマティックに解けることを期待したい。^{*5}しかし、ある意味では期待通りになるが、やはり行列という数の集まりを扱っているため少し複雑になるところがある。まず、行列には割り算という概念が存在しない。というより作れないと言った方が正しい。そのため、分数ではなく、数で言うと逆数みたいなことを行列でも考える。

数では、掛けて 1 になる数を逆数と呼んでいた。それを行列にも応用して逆数のようなものを行列でも作ってしまおうと言う話だ。まず、行列における「1」とは、どの行列にかけてもその行列の計算前と後を変えないものと考えることができる。それは単位行列と呼ばれ、2 次の正方行列では次のように定義される:

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

つまり、対角線の成分が全て 1 であり、それ以外の成分が全て 0 であるような行列を単位行列と呼ぶ。実際にどんな行列にこの単位行列をかけてもその行列を変えない性質を持っている。これを行列の世界の「1」とし、行列 A に対して、ある行列をかけると単位行列 E になるような行列を逆行列と呼ぶことにする。^{*6}行列 A を

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$$

^{*5} この表記は数学的にはよろしくない

^{*6} 一般に逆行列は正方行列のみに定義される言葉である

とすると A の逆行列は

$$-\frac{1}{3} \begin{bmatrix} -1 & -1 \\ -1 & 2 \end{bmatrix}$$

となる。実際に計算をしてみると、

$$\begin{aligned} -\frac{1}{3} \begin{bmatrix} -1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} &= -\frac{1}{3} \begin{bmatrix} -3 & 0 \\ 0 & -3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

がわかる。

では、この逆行列をどのように導出すれば良いのだろうか。それは 2 次の正方行列の場合は

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

のとき、逆行列 A^{-1} は次のように計算できる：

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

また、このときに現れた $ad - bc$ を A の行列式といい、

$$\det A$$

と書く。^{*7}

この逆行列の計算を見るとわかるとおり、 $\det A = 0$ の場合は逆行列は存在しない。この条件でない元であれば連立方程式をシステムティックに解くことが可能である。つまり、

$$Ax = b \Leftrightarrow x = A^{-1}b \quad (\det A \neq 0)$$

と連立方程式を解くことができる。

それでは、実際に連立方程式を解いてみよう。

$$\begin{cases} 2x + y = 2 \\ x - y = 4 \end{cases}$$

を解く。まずは行列に直して、

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

となる。次に、左辺の左側の行列を A 、右側の行列を x 、右辺の行列を b とおくと

$$Ax = b$$

^{*7} n 次正方行列に対する逆行列や行列式の計算方法はもっと複雑である

となる．この方程式に左から A の逆行列 A^{-1} をかけると

$$\begin{aligned} \boldsymbol{x} &= A^{-1}\boldsymbol{b} \\ &= -\frac{1}{3} \begin{bmatrix} -1 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \\ &= -\frac{1}{3} \begin{bmatrix} -6 \\ 6 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ -2 \end{bmatrix} \end{aligned}$$

のように計算ができる．もちろん，最初に提示した連立方程式の解と同じであることは言うまでもない．

このように行列を使って連立方程式を解くと，加減法などを使わずに連立方程式を解くことが可能である．これは一般の連立 n 元 1 次方程式でも同様である．方程式を行列に直し，行列の逆行列を求め，左からその逆行列を両辺に掛けて計算すると（一定の条件を満たせば）どんな連立 n 元 1 次方程式でも解ける．

5 行列の幾何的な意味

本節では，行列を図形の中でどのような意味を持つのかを述べる．簡潔に述べると行列とは，

ベクトルを他のベクトルに写す変換器

と言える．例えば，行列 A ，ベクトル \boldsymbol{x} を次のように設定する：

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

このベクトル \boldsymbol{x} を行列 A と掛けてみると，

$$\begin{aligned} A\boldsymbol{x} &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} -1 \\ 1 \end{bmatrix} \end{aligned}$$

となり，ベクトル \boldsymbol{x} が別のベクトルに変換されたことがわかる．

6 固有値と固有ベクトル

本節では，行列の固有値と固有ベクトルについて述べる．上記で行列とはベクトルを変換させるものと説明したが，ある行列に対してほぼ不変なベクトルが存在する．このほぼ不変なベクトルを固有ベクトルと呼ぶ．この固有値と固有ベクトルは線形代数学で非常に重要な役割を果たす．

6.1 固有値と固有ベクトルとは

例えば上記の A と \boldsymbol{x} は行列 A をかけてもベクトルが変化しない例である。細かく言えば変わっているが、おおらかな気持ちを持って次のようにみるとあまり変換されていないと思うのではないだろうか。

$$A\boldsymbol{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -\boldsymbol{x}$$

その他にも,

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

とすると,

$$\begin{aligned} A\boldsymbol{y} &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 3 \end{bmatrix} \\ &= 3\boldsymbol{y} \end{aligned}$$

と変換前のベクトル \boldsymbol{y} の3倍になっている。このように、行列を1つ決めるとその行列ではほとんど変わらないベクトルがいくつか存在することがわかる。

上記のようなベクトル $\boldsymbol{x}, \boldsymbol{y}$ を

固有ベクトル

と呼ぶ (正確には行列 A に対する固有ベクトル)。また、変換後のベクトルの係数のことを

固有値

と呼ぶ。一般的にまとめると次のようになる:

n 次の正方行列 A に対して,

$$A\boldsymbol{x} = \lambda\boldsymbol{x}$$

となるベクトル $\boldsymbol{x} \neq \mathbf{0}$, 定数 λ が存在するとき, \boldsymbol{x} を行列 A の固有ベクトル, λ を行列 A の固有値という。

固有ベクトルは、本来ベクトルを向きも大きさも変化させてしまう行列に対して、向きは変化せず大きさのみが変化する特別なベクトルだと思える。また λ はその変化率を指している。 λ が負の数であればそれはベクトルが真逆に向いたことを表している。

6.2 固有値と固有ベクトルの求め方

ではその固有値と固有ベクトルはどのようにして求めれば良いだろうか。上記ではいきなり登場したが、もっと複雑な固有値を持つ可能性もあり得る。これは上記で一般的にま

とめた

$$A\mathbf{x} = \lambda\mathbf{x}$$

が使える。この等式を変形して

$$(A - \lambda E)\mathbf{x} = \mathbf{0}$$

とできる。ここで E は単位行列である (行列 A と定数 λ はサイズが異なるので計算できないため λ の後ろには単位行列が隠れていると思う)。

左辺の \mathbf{x} はゼロベクトルではないため、 $A - \lambda E$ が 0 なる必要があるが、この場合は $A - \lambda\mathbf{x}$ の行列式が 0 にならなければならない。^{*8}つまり

$$\det(A - \lambda E) = 0$$

とならなければならない。行列式自体は数値なので、これは λ に関する方程式になる。よって、代数学の基本定理^{*9}により複素数の範囲で λ が求まる。

では、先ほど行列

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

の固有値を求めてみよう。まず、

$$\begin{aligned} A - \lambda E &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \\ &= \begin{bmatrix} 1 - \lambda & 2 \\ 2 & 1 - \lambda \end{bmatrix} \end{aligned}$$

なので、

$$\begin{aligned} \det(A - \lambda E) &= (1 - \lambda)^2 - 4 \\ &= \lambda^2 - 2\lambda - 3 \\ &= (\lambda + 1)(\lambda - 3) \end{aligned}$$

となる。よって、求める固有値は

$$\lambda = -1, 3 \tag{6.1}$$

である。もちろんここから固有ベクトルを求めることも可能である。

$$(A - \lambda E)\mathbf{x} = \mathbf{0}$$

となるベクトル \mathbf{x} を求めれば良い。例えば $\lambda = 3$ の場合を考える。

$$\begin{bmatrix} -2 & 2 \\ -2 & 2 \end{bmatrix} \mathbf{x} = \mathbf{0}$$

^{*8} 詳しくは線形代数の本を参照

^{*9} n 次の複素係数方程式には重複も込めてちょうど n 個の解が存在するという代数学の大定理

を満たす \boldsymbol{x} を求める． $\boldsymbol{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ とおき，行列で表示された連立方程式を展開すると

$$\begin{cases} -2x + 2y = 0 \\ -2x + 2y = 0 \end{cases}$$

となる．つまり本質的には1本の式のみで十分であることを表しているので，

$$-2x + 2y = 0$$

に注目する． y について解くと

$$y = x$$

となる．2つの文字に対して1つの式しかないので，この解を確定させること（ただ一つに決めること）は不可能である．この $y = x$ を \boldsymbol{x} に代入すると，

$$\begin{aligned} \boldsymbol{x} &= \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \begin{bmatrix} x \\ x \end{bmatrix} \\ &= x \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

が得られる．この係数の x は任意の数を取ることができ，定数倍はベクトルの向きに関係しないので，求める固有ベクトルは

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

となる．

7 行列の対角化

本節では，行列の対角化について述べる．行列の対角化とは，いわゆる行列の「標準形」を求めるお話である．行列はサイズが大きくなるほど計算といったものが大変になるのは想像に容易いと思う．なるべく簡単な形にして計算したいと思うのは数学を嗜んでいる読者なら当然考えたいことだろう．

7.1 対角行列と冪乗計算

では，行列のどの計算を簡単にしたいと思うかと聞かれれば真っ先に

行列の積

と応えるだろう．全てとまでは行かないが，特に行列の冪乗計算は行列を「標準形」にすることで容易に計算することができる．標準形とはこの資料では次のような形をした行列

を指すことにする:

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, B = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

つまり今回の「標準形」とは「対角成分」以外全て 0 となるような行列を指す. このような行列を一般に対角行列という. この対角行列は冪乗計算と非常に相性がいい. 例えば A の冪乗を計算してみる.

$$\begin{aligned} A^2 &= \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}, \\ A^3 &= A^2 A \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 8 \end{bmatrix} \end{aligned}$$

となる. 一般に自然数 n に対して A^n は次のようになる:

$$A^n = \begin{bmatrix} 1^n & 0 \\ 0 & 2^n \end{bmatrix}$$

このように行列は対角行列であれば, 冪乗の計算が容易になる.

7.2 行列の対角化

つまり行列の対角化とは, 与えられた行列を上記のような対角行列に変換することを指す. では与えられた行列をどのように対角化することができるのかという問いに関しては前節の固有値と固有ベクトルが密接に関係しておりそれらを用いて対角化することが可能である. 対角化は, 行列にある行列を掛け算することで実現される. 例を見ることにする.

行列 A を

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

とし, 行列 P を

$$P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

とする. このとき, $P^{-1}AP$ を計算する.

$$P^{-1} = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

なので,

$$\begin{aligned} P^{-1}AP &= \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} -2 & 0 \\ 0 & 6 \end{bmatrix} \\ &= \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \end{aligned}$$

のように対角行列に変換することができる. ここに現れる対角成分 $-1, 3$ は (6.1) で現れた固有値に等しいがこれは偶然ではない. このように行列の対角化では, 対角化しようとする行列の固有値が対角成分となり, 対角化をすることができる. また, 行列 P についても固有ベクトル \mathbf{x}, \mathbf{y} になっていることもわかる. もちろんこれも偶然ではなく, 対角化するための行列 P は固有ベクトルを並べて作られる (今回は $P = [\mathbf{x} \ \mathbf{y}]$ のように並べた).

上記のような固有値と固有ベクトルを用いて対角化できる原理は次のとおりであり, 以下の説明は一般に n 次正方行列でも同様である.

\mathbf{x}, \mathbf{y} を固有ベクトル, それぞれの固有値を λ_x, λ_y とする. このとき,

$$A\mathbf{x} = \lambda_x\mathbf{x}, \quad A\mathbf{y} = \lambda_y\mathbf{y}$$

が成り立つ. これらを次のように行列にする:

$$[A\mathbf{x} \quad A\mathbf{y}] = [\lambda_x\mathbf{x} \quad \lambda_y\mathbf{y}]$$

さらに,

$$\begin{aligned} [A\mathbf{x} \quad A\mathbf{y}] &= A [\mathbf{x} \quad \mathbf{y}], \\ [\lambda_x\mathbf{x} \quad \lambda_y\mathbf{y}] &= \begin{bmatrix} \lambda_x & 0 \\ 0 & \lambda_y \end{bmatrix} [\mathbf{x} \quad \mathbf{y}] \end{aligned}$$

と変形できるので (A や $\mathbf{x}, \mathbf{y}, \lambda_x, \lambda_y$ を具体的なものに置き換えて計算してみると計算の意味がわかる. また, ベクトルや行列は 1 つの数のように考えてみる),

$$P = [\mathbf{x} \quad \mathbf{y}]$$

とおけば,

$$AP = \begin{bmatrix} \lambda_x & 0 \\ 0 & \lambda_y \end{bmatrix} P$$

より,

$$P^{-1}AP = \begin{bmatrix} \lambda_x & 0 \\ 0 & \lambda_y \end{bmatrix}$$

が成り立つ.*10

*10 行列 P は必ず逆行列をもつことが証明されている.

以上が対角化の原理である．行列 A の固有値が対角行列の対角成分となり，対角化する行列 P は固有ベクトルを並べたものであるとわかる（固有値と固有ベクトルの順番は合わせる必要がある）．

7.3 行列の対角化の応用

行列の対角化は冪乗計算において絶大な威力を発揮する．次の行列の n 乗を計算することを考える．

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

2 乗，3 乗と計算し，規則を見つけて n 乗の場合を推測して数学的帰納法で示すことも可能だが，少し博打がすぎるとも思える．行列の対角化をすることで行列の n 乗計算を楽にできる．*¹¹前小節において，行列 A は 2 行 2 列の行列

$$P = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

を用いて，

$$P^{-1}AP = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \quad (7.1)$$

とできた．この両辺を n 乗することを考える．右辺は対角行列なので，

$$\begin{bmatrix} (-1)^n & 0 \\ 0 & 3^n \end{bmatrix}$$

が成り立つ．ここで，左辺の n 乗について考える．まずは 2 乗について考えると

$$\begin{aligned} (P^{-1}AP)^2 &= (P^{-1}AP)(P^{-1}AP) \\ &= P^{-1}A(P P^{-1})AP \\ &= P^{-1}AEAP \\ &= P^{-1}A^2P \end{aligned}$$

となる．3 乗すると（計算は省略するが）

$$(P^{-1}AP)^3 = P^{-1}A^3P$$

となる． P と P^{-1} がうまく打ち消しあって中の A のみ冪乗計算されることになることがわかる．よって， n 乗でも同様に

$$(P^{-1}AP)^n = P^{-1}A^nP$$

が成り立つ．よって，

$$P^{-1}A^nP = \begin{bmatrix} (-1)^n & 0 \\ 0 & 3^n \end{bmatrix}$$

*¹¹ 多項式の割り算を用いた行列の n 乗計算も存在する

となる．この式の左から P ，右から P^{-1} をかけることで

$$\begin{aligned} A^n &= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} (-1)^n & 0 \\ 0 & 3^n \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} (-1)^n & -(-1)^n \\ 3^n & 3^n \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} (-1)^n + 3^n & -(-1)^n + 3^n \\ -(-1)^n + 3^n & (-1)^n + 3^n \end{bmatrix} \end{aligned}$$

が成り立つ．確認として $n = 1$ を代入すると，

$$\begin{aligned} A^1 &= \frac{1}{2} \begin{bmatrix} (-1) + 3 & -(-1) + 3 \\ -(-1) + 3 & (-1) + 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \end{aligned}$$

となり，確かに正しいことが確認できる．

8 行列の特異値分解

本節では行列の特異値分解について述べる．機械学習分野においては非常によく登場する理論である．行列の本質的な部分を抽出するときに対角化や特異値分解が強力な役割を果たす．

8.1 特異値分解とは

前節までの行列の対角化は全て正方行列に限られた話であったが，やはり正方行列ではない一般のサイズの行列に対しても対角化のような操作を行いたいと考えるのは当然の発想である．それが特異値分解と呼ばれるもので，対角化の一般化として知られる．次の行列 A は正方行列ではないため，対角化をすることができない．

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$$

しかし，特異値分解であれば可能である．^{*12}

8.2 特異値分解の計算方法 (概略)

特異値分解の計算手順は多く，煩雑なものになるのでステップ分けして概要を述べる．ここで，下記に現れる tA とは転置行列を表し，行列 A の (i, j) 成分を (j, i) 成分にした行列である．行と列のサイズが入れ替わる．

Step 1 tAA , $A{}^tA$ の固有値を求める．

^{*12} 任意の行列に対して特異値分解可能であることが証明されている．

まず, tAA , A^tA はそれぞれともに正方行列になることに注意する. そして, 実は tAA , A^tA の 0 でない固有値はともに同じであり, その値は 11, 1 である (3×3 行列の方からは 0 の固有値が現れる).

Step 2 tAA , A^tA のそれぞれの固有ベクトルを求める.

tAA は 2×2 行列なので, (大きさを 1 にした) 固有ベクトルは 2 つありそれぞれ

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

となる. また, A^tA は 3×3 行列なので (大きさを 1 にした) 固有ベクトルは 3 つありそれぞれ

$$\mathbf{u}_1 = \frac{1}{\sqrt{22}} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{u}_3 = \frac{1}{\sqrt{11}} \begin{bmatrix} 1 \\ -3 \\ 1 \end{bmatrix}$$

となる. この $\{\mathbf{v}_1, \mathbf{v}_2\}$ と $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ はそれぞれ互いに直交しており, 大きさも 1 であることに注意する.

Step 3 直交行列 V, U を求めて特異値分解する

$$V = [\mathbf{v}_1 \quad \mathbf{v}_2], \\ U = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3]$$

とすると V, U は直交行列になる.*¹³

よって,

$tVAU$

を計算すると

$${}^tVAU = \begin{bmatrix} \sqrt{11} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

が成り立つ.

これが特異値分解である.*¹⁴機械学習分野で扱われる行列は一般に行数と列数が異なる場合が多いので, そのような場合でも対角化を行い計算を楽にしたり, 行列の本質的な重要な部分を抜き出す操作としてこの特異値分解が利用される.*¹⁵

9 おまけ: 行列の Jordan 標準形

最後の節では, 行列の Jordan 標準形について少しだけ述べる.

*¹³ 転置行列が逆行列になるような行列のこと. つまり ${}^tA = A^{-1}$ となる行列 A のことであり, 良い性質を多くもつ

*¹⁴ 正確には $A = V\Sigma^tU$ と行列 A を直交行列と対角行列の積に分解することを特異値分解と呼ぶ

*¹⁵ 例えば教師なし学習における次元削減など

Jordan 標準形とは、行列の対角化の一般化であり、全ての正方行列は Jordan 標準形と呼ばれる行列に変換することができる。

例えば、次の行列は対角化することができない。

$$A = \begin{bmatrix} 4 & -1 \\ 1 & 2 \end{bmatrix}$$

実際、固有値を求めると $\lambda = 3$ のみで、固有ベクトルも $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ のみである。 A の対角化には固有値と固有ベクトルが 2 つ必要であるがここでは 1 つしか取ることができない。このように $\det(A - \lambda E) = 0$ が重解をもつ場合には対角化できないこともある。

では対角化を諦めるのかということそうではない。なるべく対角化と似たような変換ができないかを考える。それこそが Jordan 標準形であり、任意の正方行列は Jordan 標準形に変換できることが証明されている。詳しい理論や計算方法を述べることはここではしないが、例えば行列 A は次のような Jordan 標準形へと変換される。

$$J = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$$

このように、Jordan 標準形においても対角成分に固有値が並ぶ。対角行列とまではいかないが、かなり近い形に変換されていることがわかる。また、行列 J の (1,2) 成分の 1 は偶然ではなく、Jordan 標準形に変換すると対角成分の上に 1 がいくつか乗る。Jordan 標準形に変換することで冪乗計算や、複雑な行列でもこの形に変換することで解析が可能となり、さまざまな行列の性質を調べることができる。