きらり。中のヒト 第9回

やんわり理解する行列のランク

松井勇佑†

まえがき

東京大学にてコンピュータビジョンの研究を行っている松井勇佑です．自分は今でこそ大学教員として数学やプログラミングを学生に教える立場にありますが，振り返ると大学入学当初の線形代数にはまったくついていけませんでした．その後卒論を書く段階になり，ようやく線形代数というものが少しずつわかるようになり，その重要性を認識しました．何より，線形代数とは実際に現実の問題を解く際に必要な「実学」である，ということがわかりました．

本稿では，なぜ入学当初に線形代数を難しく感じたのかをまず振り返ります．そして今だからこそなんとなく持っている行列計算の「フィーリング」を，ベクトルの外積から始めて説明してみようと思います．証明はゼロです．不正確な記述だらけですが，あくまでフィーリングを伝えるものだと思ってお許しください．もし線形代数に対して苦手な気持ちを持っている方がいましたら，本稿が少しでも理解の助けになりましたら幸いです．

挫折の大学1年生

私は大学1年生の線形代数の講義にまったくついていけませんでした．今でも思い出すのは掃き出し法です．掃き出し法とは，「行列$\boldsymbol{A}$を考える．$\boldsymbol{A}$の1行目を定数倍したものを2行目から引く．このような操作を繰り返すことで行列を階段の形状にできる」というものです．そして，得られた階段のでっぱりの部分の個数が「その行列のランクである」というものです．

この説明を聞いたとき，私は完全にチンプンカンプンになってしまいました．まず，① なぜそのような作業をしていいのかわからない，そして，② その結果得られる階段にどういう意味があるのかわからない，という疑問が沸きました．ここで，①については，$\boldsymbol{Ax}=\boldsymbol{b}$を考えるときに，行を足したり引いたりする操作に相当する行列$\boldsymbol{E}$（基本行列）を両辺にかけて$\boldsymbol{A'x}=\boldsymbol{b'}$を考える（ここで$\boldsymbol{EA}=\boldsymbol{A'}$，また$\boldsymbol{Eb}=\boldsymbol{b'}$）というのがやりたいことなんだなということはのちに理解しました．しかし，②の疑問，すなわち「そもそもこれはどういうことなんだろう」という思想レベルの問題理解がまったくできませんでした．

もちろん，知識や手続きとしては，左辺を階段状にすれば変数を順番に下から消去していくことで方程式を解けるということはわかります．しかし，ここでいう「ランク」の意味する「お気持ち」が，不明でした．今から考えると，このころは実際にデータに触れることなしに証明だけを追っていたので，お気持ちが理解できないのは当然だったのかもしれません．私を含め，多くの工学系の学生は，実際にデータに触れることで，このお気持ちが理解できる可能性があると感じます．もし今1年生で線形代数に悩んでいる学生がいましたら，実際にデータに触れる卒論まで少し我慢して勉学を続けてみることをオススメします．

ランクとは情報の詰まり具合

さて，それでは一体ランクとは何を意味するのでしょうか？ それは，「行列にデータがどれほどつまっているか」を表す指標といえば直感的ではないか，というのが最近の私の意見です．以下に，ランク1から始めて，順番に考えていきましょう．

まず例として縦ベクトル$\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^{3\times1}$を考えましょう．

$$\boldsymbol{v}_1 = \begin{bmatrix} 6 \\ 3 \\ 5 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} 9 \\ 2 \\ 8 \end{bmatrix}.$$

ここで，ベクトルの外積は，行列を形成します．これを$\boldsymbol{A}$としましょう．

$$\boldsymbol{A} = \boldsymbol{v}_1\boldsymbol{v}_2^\top = \begin{bmatrix} 54 & 12 & 48 \\ 27 & 6 & 24 \\ 45 & 10 & 40 \end{bmatrix}.$$

さて，この行列をよく見ると，独特な繰り返し構造をもっていることに気付きます．例えば1行目は2行目の2倍になっています．3列目は2列目の4倍です．これは当たり前で，このような繰り返しは，外積の計算の定義からの帰結そのものだからです．結局この行列はわずか1本の列ベクトル（$\boldsymbol{v}_1$）の定数倍（9倍，2倍，8倍）を並べたものとして表現されます．この係数を並べたものが$\boldsymbol{v}_2$そのものです．

†東京大学
"Understanding the Rank of a Matrix Softly" by Yusuke Matsui (The University of Tokyo, Tokyo)

映像情報メディア学会誌 Vol. 77, No. 1, pp. 62～64 (2023)

これは何を意味するのでしょうか？行列$\boldsymbol{A}$は，本来は3×3の9個の数字を使わないと表現できないはずだったのに，ここではたった2本のベクトル$\boldsymbol{v}_1, \boldsymbol{v}_2$で表現できてしまっています．すなわち，行列内部に繰り返し（定数倍の足し引き）が多く，この行列で表現される情報は冗長であるという気になってきます．言い換えると，$\boldsymbol{A}$には情報があまり詰まっていない気持ちになります．これが，ランク1の定義です．

$\mathrm{rank}(\boldsymbol{A}) = 1.$

すなわち，「行列がベクトルの外積で表現できてしまうとき，その行列は情報をあまりもっておらず，ランクが1」です．あるいは，「行列の各行または各列を定数倍の差し引きの意味で整理したときに，1本のベクトルと係数ベクトルで表現できるとき，ランクが1」と言ってもいいです．実際，行列をベクトルの外積で無理やり表現することをrank-1近似といったりします．線形代数の教科書ではこのランク1の話を明示的にすることはあまりない，あるいはあったとしても後半の応用パートであることが多い気がするのですが，個人的にはこの「ベクトルの外積＝ランク1の行列」という話を最初にすると見通しがよいのではと考えます．

ちなみに，行ベクトル（高さが1の行列）および列ベクトル（幅が1の行列）のランクは1です．これは「情報の詰まり具合がランクである」から直感的に考えて，ベクトルの場合はとにかくランクが1だと考えればよいです（注意として，全要素が0の場合のみ，ランクは0です）．

さて，それでは次に「ベクトル（3×1の行列）ではなく，3×2の行列である$\boldsymbol{v}_3, \boldsymbol{v}_4 \in \mathbb{R}^{3\times 2}$」，およびそれに対し$\boldsymbol{v}_4$を転置し積をとった$\boldsymbol{B}$を考えましょう．

$$\boldsymbol{v}_3 = \begin{bmatrix} 19 & 30 \\ 15 & 23 \\ 8 & 18 \end{bmatrix}, \quad \boldsymbol{v}_4 = \begin{bmatrix} 29 & 27 \\ 10 & 27 \\ 9 & 26 \end{bmatrix}.$$

$$\boldsymbol{B} = \boldsymbol{v}_3 \boldsymbol{v}_4^{\top} = \begin{bmatrix} 1361 & 1000 & 951 \\ 1056 & 771 & 733 \\ 718 & 566 & 540 \end{bmatrix}.$$

この計算は，最初の「ベクトル外積」を少し一般化したものになっています．すなわち，最初のケースでは2本のベクトル（3×1）を考えましたが，ここでは少し幅が広い3×2の行列を考え，その積をとっています．ここでは，想像がつく通り，$\boldsymbol{B}$のランクは2になります．

$\mathrm{rank}(\boldsymbol{B}) = 2.$

この行列$\boldsymbol{B}$は，幅2の細長い行列の積で表現できてしまっています．なので，やはり情報はあまり詰まっていません．ですが幅1のときよりは詰まっています．これを表現するのが，ランクという概念です．ここでは紙面の都合で3x3の行列を用いていますが，もしこれが100x100の行列だった場合，$\boldsymbol{B}$もまた細長い行列2本で表現されてしまうということがわかると思います．

ここで，幅2の行列$\boldsymbol{v}_3, \boldsymbol{v}_4$もランクが2です．これは「情報の詰まり具合がランク．幅1の行列であるベクトルはランク1．なので幅2の行列の場合には普通はランク2」とうことです．ちなみに行列が細長いとき，ランクがとりうる最大値は，行列の短いほうの長さになるので，この場合では最大でも2です．

ちなみに上で「普通は」とか「最大でも」と言っているのは，そうでない場合があるからです．次を考えましょう．

$$\boldsymbol{v}_3 = \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 10 & 20 \end{bmatrix}$$

上のようにもし元の$\boldsymbol{v}_3$が「各列が別の列の定数倍」のようになっている場合，これはランク1になってしまうので，$\boldsymbol{B}$もランク2になりません．これは当然ながら，「情報の詰まり具合」というランクの定義からいうと，この$\boldsymbol{v}_3$にはあまり情報が詰まっていない（各列を定数倍して差し引きして整理すると，1本の列ベクトルで表現できてしまう）からです．

ここまでをまとめておきます．

(1) 行列のランクとは，行列に「情報がどれだけ詰まっているか」を表す指標．
(2) ここで詰まり具合というのは，各行または各列に対し定数倍の足し引きを施し整理できるかどうかというニュアンス．
(3) ベクトルのランクは1．
(4) 行列がベクトル2本の外積で表現できるとき，その行列のランクも1．
(5) ちょっと幅広（ランク2）のベクトルの積で行列が表現できるとき，その行列のランクも2．
(6) もっと幅広（ランクk）のベクトルの積で行列が表現できるとき，その行列のランクもk．

よって，繰り返しになりますが，行列のランクというのは，その行列がもつ情報の量を表す概念です．例え行列が巨大でも，そのランクが小さい場合は，すごく細長い行列の積で表現できてしまいます．なので，実質的に元の巨大な行列が持つ情報は少ないということです．

データ処理での実例

さて，上記の話は実際にデータ処理を行う上で極めて直感的に理解されます．データ処理を行う上で，行列は「観測やデータを並べたもの」を表現するために使われることが多いです．ここではオンラインショッピングサイトにてお客さんの購買記録を分析することを考えましょう．1人の購買記録を1本の行ベクトルとして表します．各要素は，「Xというブランドが好きだというスコア」とか「Yというブランドを買った回数」とかだとしましょう．この行ベクトルを並べて巨大な行列$\boldsymbol{C}$を作ります．例えば以下のような形です．

$$\boldsymbol{C} = \begin{bmatrix} 3.2 & 13 & 7 & & 4 \\ 4 & 8.5 & 91 & \cdots & 88 \\ & \vdots & & & \\ 13 & 22.1 & 8 & & 3.7 \end{bmatrix}$$

この行列は横にとても長いとしましょう．ここでもし各行が互いに全然違う場合は，多種多様なユーザによる購買記録が集まっていると言えます．したがって，この$\boldsymbol{C}$を分析することで何らかの知見を得られそうです．これまで

の議論から明らかなようにその場合は「各行が，他の行の線形和で表現できない」ため$\boldsymbol{C}$のランクは大きくなります．

一方で，ここでもし以下のように，「すごく似ている行がたくさんある」場合はどうでしょうか？

$$\boldsymbol{C}=\begin{bmatrix} 3.2 & 13 & 7 & & 4 \\ 3.1 & 12.9 & 7 & \dots & 4.2 \\ & & \vdots & & \\ 13 & 22.1 & 8 & & 3.7 \end{bmatrix}$$

ここでは，1行目と2行目が似ています．これは，ユーザ1とユーザ2による購買記録が似ていることを意味します．もしそのようなケースがたくさんあった場合，このデータマトリクス$\boldsymbol{C}$は巨大でも，実質的にはほんの少しの典型ユーザしかいない可能性があります．残りのユーザは，その典型ユーザに似たユーザです．その場合，この$\boldsymbol{C}$を分析してもあまりいい知見が得られないかもしれません．例えば，ひょっとしたらボットアカウントが生成した記録が大量に含まれているかもしれません．

この場合，「各行は，他の行の線形和でかなりうまく近似できる」ということなので，行列$\boldsymbol{C}$は低ランク行列$\boldsymbol{D}$でうまく近似できます．すなわち，

$$\mathrm{rank}(\boldsymbol{D})=小$$

である$\boldsymbol{D}$を用いて$\|\boldsymbol{C}-\boldsymbol{D}\|$をかなり小さくできます．ここで差異の指標は何らかのノルムです．そして，$\boldsymbol{D}$は低ランクなので細長い行列の積で表現できます．上記の議論はかなりいい加減なものです．しかし，これはAmazonなどのEコマースサイトにおいて，過去の購入履歴からユーザが次に買うであろうものを推薦する「推薦システム」を構築する際の，数学的な基盤となります．

このように，ランクという概念は，データ行列にどれくらい情報がつまっているかを表す直感的な指標であり，データを扱う際に直接的に使えます．「この行列はランク落ちだね」とかいうフレーズは，「本来はもっとランクが高くあってほしい，すなわち各行が多様であってほしいが，いま対象にしている行列はそれほど多様ではない」という意味になります．例えば，私の分野ではざっくり言うと「ある位置からカメラで撮影した画像」が一行を構成するデータ行列を作ったりしますが，そこではランク落ちとは「変な位置から撮影してしまったため，得られる情報が少ない（ill-pose）」といった，物理的な意味も持ったりします．

振り返って掃き出し法

それではここで振り返って，掃き出し法とはなんだったのでしょうか？ここまでの話をもとにすると，実は掃き出し法は「ランクを手作業でチェックする」行為であるととらえることができます．行列$\boldsymbol{A}$を思い出しましょう．これに各行を消去する基本行列を左からかけてみると，

$$\begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -45/54 & 0 & 1 \end{bmatrix}\boldsymbol{A}=\begin{bmatrix} 54 & 12 & 48 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

となり，階段のでっぱりが一つなので，$\boldsymbol{A}$のランクは1であるとわかります．これは，これまでさんざん述べてきた「各行・各列について定数倍で整理できる」といっている行為そのものです．

行列$\boldsymbol{B}$についても考えてみましょう．まず第1行の定数倍を各行から引いた$\boldsymbol{B}'$を計算します．

$$\begin{bmatrix} 1 & 0 & 0 \\ -1056/1361 & 1 & 0 \\ -718/1361 & 0 & 1 \end{bmatrix}\boldsymbol{B}=\begin{bmatrix} 1361 & 1000 & 951 \\ 0 & -4.9 & -4.88 \\ 0 & 38.45 & 38.3 \end{bmatrix}=\boldsymbol{B}'.$$

次に$\boldsymbol{B}'$の2行目の定数倍を引いて3行目を消します．

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 38.45/4.9 & 1 \end{bmatrix}\boldsymbol{B}'=\begin{bmatrix} 1361 & 1000 & 951 \\ 0 & -4.9 & -4.88 \\ 0 & 0 & 0 \end{bmatrix}$$

よって，階段のでっぱりは2ヵ所なので，$\boldsymbol{B}$のランクは2だということがわかります．

このように，掃き出し法のお気持ちは，「行列に対する情報の詰まり具合を示すランク」を手計算で得ていたことに相当する，と理解することができます．

むすび

大学1年生のころはチンプンカンプンだった掃き出し法とランクですが，データを実際に扱うようになるとお気持ちがなんとなくわかるようになりました．今になって線形代数の教科書[1) 2)]を読むと，なんて面白いんだという感想を持ちます．座学で得た知識は，実際に使ううちになんとなくわかってくるものですね．

（2022年10月25日受付）

〔文　献〕

1）ギルバートストラング：“線形代数とその応用”，産業図書（1978）
2）室田一雄，杉原正顯：“基礎系数学線形代数Ⅰ（東京大学工学教程）”，丸善出版（2015）

《付　録》

外積がrank-1行列を形成することを示すソースコードの例（juliaをインストールしてください）．

```
$ julia

julia> using LinearAlgebra

julia> v1 = rand(1:200, 3, 1)
3 x1 Matrix{Int64}:
86
163
66

julia> v2 = rand(1:200, 3, 1)
3x1 Matrix{Int64}:
109
180
84

julia> A = v1 * v2'
3x3 Matrix{Int64}:
9374 15480 7224
17767 29340 13692
7194 11880 5544

julia> rank(A)
1
```