データ行列の基本構造 *

Susan C. Weller, A.Kimball Romney
訳：藤本一男 †

2016-04-03

計量尺度法の中心的操作は、データ行列を基本構造へと分解することである。その分解は、元の変数の要素に対応した情報を体現する 3 つの行列の積として表現される。行列を基本構造に分解するプロセスは、特異値分解 (SVD[*1]) と呼ばれる。本モノグラフの目的からすると、我々は、SVD をブラック・ボックスとして扱い、その導出過程にではなく、分解された要素行列の解釈と応用に焦点をあてる。数理的アルゴリズムとソフトウェアが、この過程を遂行するために提供されており、もし読者がその導出過程やアプリケーション・パッケージについて説明を必要とするならば、関連する外部情報を Appendix で提供した。本章においては、データ行列の基本構造を説明し、以後の章で用いられる正規化処理 (normarization procedure) を整理しておく。読者は、統計学の基礎知識と行列代数の簡単な演算について慣れていることを想定されている。

1　行列の基本構造

データの基本構造の描写から始めよう。我々は、データのセットを、つまり、データ行列を X として参照する。そして、X のある観測値を x_{ij} として参照する。この行列は、各行が個別の主体を表し、各列がそれぞれのテスト項目に対応しているような、個人のグループに対する心理学的テストの得点かもしれない。行列の各セル x_{ij} は、テスト j に対する、主体 i の得点を体現している。もしくは、行列は、そのセルの値が、対応する実際の度数になるような分割表を体現するものでもよい。例えば、行変数は「都市」で、列が「犯罪の種類」を表わし、各セルは、各都市で発生する犯罪タイプの数を示す、というようにである。

すべてのデータ行列は、「特性」要素へと分解することができる。n 行 m 列の行列 (n>m) は、基本構造情報としては、例 2.1 に示されているように、3 つの行列から構成される。

行列 U は、X の行における情報を「要約」している。U の行は、X における行に対応し、U の列は、行変数における背後次元、要素を表現している。同様に、V の行は、X の列に対応し、V の列は、X の列の潜在要素を要約している。U の行数と X の行数は、等しく、V の行数は、X の列数に

---

* 本稿は、*METRIC SCALING* ,SAGE Publications 1990,「計量尺度法」第 2 章の試訳である。

† kazuo.fujimoto2007@gmail.com

[*1] Singular Value Decompression

等しい。U と V の行列の列は、データ構造の潜在次元、もしくは、基本要素 (basic components) を表現している。

d 行列は、特別な種類の行列で、対角行列であり、対角要素以外がすべてゼロの正方行列である。d における対角要素は、U,V 行列の列に対応する特異値からなっている。最初の要素の d_{11} は、U の最初の列、そして V の最初の列に対応し、第 2 の要素である d_{22} は、U の第 2 列、V の第 2 列に対応している。d の値は、U と V の次元の相対的な「重要性」の「重み」を表しており、大きなものから小さいものの順にならんでいる。U と V の列と d の要素は、X の構造全体で、重要なものからより重要でないものへという順にならんでいる。もし、X に冗長情報がないのであれば、U と V の列の数と d の次元は、X の少ない方法次元する m に等しい。

例 2.2 は、5×3 行列での数値例を示している。X の基本構造が、U, d, V 行列によって表現されている。実際のところ、X は、U、d、V^T 行列の積である。V の転置 (V^T) が、行列の計算で用いられることに注意すること。

$$X_{(n\times m)} = U_{(n\times m)} d_{(m\times m)} V^T_{(m\times m)} = U d V^T_{(n\times m)}$$

行列の基本構造は、玉ねぎの皮のようなものであって、要素を一枚一枚剥いていくことができる。そして、部分、もしくは、全体が似ている。X の 1 次元近似は、U の第 1 列ベクトル、d の最初の要素、V の第 1 列 (V^T の第 1 行) をかけあわせることによって得られる。2 次元近似は、U の最初の二つの列ベクトル、d の最初から二つの要素、V の二つの列ベクトルをかけあわせることで得られる。X の三次元近似は、U の列ベクトルの三つすべて、d の要素三つ、V の三つの列ベクトルを用いて得ることができる。

例 2.1
基本構造の表記法

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix} \rightarrow \begin{bmatrix} U_{11} & U_{12} & \dots & U_{1m} \\ U_{21} & U_{22} & \dots & U_{2m} \\ \vdots & \vdots & & \vdots \\ U_{n1} & U_{n2} & \dots & U_{nm} \end{bmatrix} \begin{bmatrix} d_{11} & 0 & \cdot\cdot & 0 \\ 0 & d_{22} & \cdot\cdot & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdot\cdot & d_{mm} \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} & \cdot\cdot & V_{1m} \\ V_{21} & V_{22} & \cdots\cdot & V_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ V_{n1} & V_{n2} & \cdot\cdot & V_{mm} \end{bmatrix}$$

例 2.2 に示す、X に対するこうした段階的な近似を例 2.3 にしてしてある。X の 1 次元推定は、行列の水平セットの最初のものとして図示される。第 2 のセットは、X を第 2 要素からのみ推定する。X の最初の要素 (U_{i1},d_{11} ,V^T_{1j}) からの推定と、第 2 の要素 (U_{i2} ,d_{22} ,V^T_{2j}) からの推定は、それぞれ、セルごとに加算され、一番下の 2 次元推定を得ることになる。同様にして、1 次元の、2 次元の、そして 3 次元の推定を得ることができる。

例 2.2
基本構造の例

$$\begin{bmatrix} 2 & 8 & 10 \\ 5 & 3 & 1 \\ 4 & 9 & 15 \\ 20 & 10 & 5 \\ 15 & 18 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} .287 & -.477 & -.066 \\ .144 & .152 & .029 \\ .398 & -.641 & .454 \\ .557 & .577 & .535 \\ .654 & .075 & -.709 \end{bmatrix} \begin{bmatrix} 37.948 & 0 & 0 \\ 0 & 14.700 & 0 \\ 0 & 0 & 4.888 \end{bmatrix} \begin{bmatrix} .628 & .674 & .389 \\ .623 & -.137 & -.770 \\ .465 & -.726 & .506 \end{bmatrix}$$

もし、X が冗長情報を含んでいると、それは、U, d, V の中に集約される。例 2.4 は、例 2.2 に 3 列を追加したものである。第 4 の列は、第 1 の列のまったくのコピーであり、第 5 列は、第 2 列の 2 倍であり、第 6 列は、第 2、3 列を加算したものになっている。言い換えると、新たな行列には、新しい情報はない。そして、新たな列のデータは、最初の 3 つの列と完全に冗長である。V において、第 1、第 4 行は、同一である; そして、第 5 行に対する座標は、第 2 行の 2 倍になっているということ。第 6 行の座標は、第 2、3 行の和の座標になっているということを確認して欲しい。こうして、X の有意味な次元は 3 次元ということになる。この最大次元は、行列 X の $Rank$ とも呼 ばれる。

例 2.3

次元縮減によるデータ推定

1 次元推定

$$\begin{bmatrix} .287 \\ .144 \\ .398 \\ .557 \\ .654 \end{bmatrix} \begin{bmatrix} 37.948 \end{bmatrix} \begin{bmatrix} .628 & .623 & .465 \end{bmatrix} = \begin{bmatrix} 6.847 & 6.795 & 5.072 \\ 3.441 & 3.415 & 2.549 \\ 9.489 & 9.418 & 7.030 \\ 13.272 & 13.172 & 9.832 \\ 15.602 & 15.484 & 11.558 \end{bmatrix}$$

$$U_{i1} \quad d_{11} \quad V_{1j}^T \quad = \quad U_{i1} d_{11} V_{1j}^T$$

$$\begin{bmatrix} -.477 \\ .152 \\ -.641 \\ .577 \\ .075 \end{bmatrix} \begin{bmatrix} 14.700 \end{bmatrix} \begin{bmatrix} .674 & -.137 & -.726 \end{bmatrix} = \begin{bmatrix} -4.721 & 0.958 & 5.090 \\ 1.504 & -0.305 & -1.611 \\ -6.352 & 1.288 & 6.849 \\ 5.710 & -1.158 & -6.156 \\ 0.747 & -1.151 & -0.805 \end{bmatrix}$$

$$U_{i2} \quad d_{22} \quad V_{2j}^T \quad = \quad U_{i2} d_{22} V_{2j}^T$$

2 次元推定

$$\begin{bmatrix} .287 & -.477 \\ .144 & .152 \\ .398 & -.641 \\ .557 & .577 \\ .654 & .075 \end{bmatrix} \begin{bmatrix} 37.948 & 0 \\ 0 & 14.700 \end{bmatrix} \begin{bmatrix} .628 & .623 & .465 \\ .674 & -.137 & -.726 \end{bmatrix} = \begin{bmatrix} 2.125 & 7.753 & 10.162 \\ 4.944 & 3.110 & 0.928 \\ 3.137 & 10.707 & 13.879 \\ 18.982 & 12.014 & 3.676 \\ 16.349 & 15.333 & 10.753 \end{bmatrix}$$

$$U_2 \qquad d_2 \qquad V_2^T \qquad = \qquad U_2 d_2 V_2^T$$

ランク ($Rank$) は、データ・ポイントの布置によって測られる下位空間の次元である。これはまた、d における非ゼロ要素の数に等しい。例 2.4 において、X の各行、6 次元空間に位置すると考えることができる。しかし、d における要素数は、この布置が情報のロスなしで、3 次元で表現できるということを示している。ランクを K として、その次元で布置を表現することができる。

$$X_{(n\times m)} = U_{(n\times k)} d_{(k\times k)} V_{(m\times k)}^T$$

こうして、行変数は、元の 6 次元ではなく 3 次元で記述されることになる。

例 2.4

縮減された $Rank$ による基本構造

$$\begin{bmatrix} 2 & 8 & 10 & 2 & 16 & 18 \\ 5 & 3 & 1 & 5 & 6 & 4 \\ 4 & 9 & 15 & 4 & 18 & 24 \\ 20 & 10 & 5 & 20 & 20 & 15 \\ 15 & 18 & 9 & 15 & 36 & 27 \end{bmatrix} \rightarrow \begin{bmatrix} .331 & .452 & .006 \\ .129 & -.166 & -.023 \\ .423 & .561 & -.531 \\ .470 & -.669 & -.510 \\ .688 & -.074 & .676 \end{bmatrix} \begin{bmatrix} 77.585 & 0 & 0 \\ 0 & 21.273 & 0 \\ 0 & 0 & 7.748 \end{bmatrix} \begin{bmatrix} .293 & -.572 & -.294 \\ .308 & .007 & .294 \\ .236 & .412 & -.566 \\ .293 & -.572 & -.294 \\ .617 & .013 & .589 \\ .545 & .418 & -.272 \end{bmatrix}$$

$$X_{(5\times 6)} \qquad U_{(5\times 3)} \qquad d_{(3\times 3)} \qquad V_{(6\times 3)}$$

基本構造行列は、空間的には、n 次元ユークリッド空間における X のデータ・ポイントを、再表示もしくは「回転」したものである。U, d そして V 行列は、データ・ポイントを新たな座標軸に「再配置」する。元のデータ・ポイントの配置が回転され、延伸され、もしくは縮尺され、反転 (ミラー・イメージ) されることもある。そして、X の行変数は、U の列座標となり、X の列変数は V の列座標となる。座標ベクトルは、お互いに直行しており、単位長さに標準化されている。

$$1.0 \quad = \quad (\sum_i U_{ik}^2)^{1/2} \quad = \quad (\sum_i V_{ik}^2)^{1/2}$$

もし、行列が対称行列であるならば、行と列の要素行列は等しいものとなる。正方行列 S の基本構造は、次のようになる。

$$S = UdVt = UdU^T = VdV^T$$

転置行列を、pre、もしくは post に乗算することは (XT^X、XX^T) は、正方、対称行列を得る。X、XX^T、もしくは、X^TX の分解は、「同じ」基本構造となる。

$$X = UdV^T$$

$$XX^T = Ud^2U^T$$

$$X^T X = V d^2 V^T$$

さらに、$rank(X) = rank(X^T) = rank(XX^T) = rank(X^T X)$ である。

正方、対称行列に対する基本構造は、固有構造 (eigen structure) に等しい (eigen は、ドイツ語で「特性」を意味する)。この特別なケースにおいて、対称行列の分解は、U、V という構成列ベクトルは、固有ベクトルもしくは特性ベクトルとして参照される。d の対角要素、つまり特異値は、固有値、特性根もしくは潜在根とも呼ばれる。対称行列の分解のいま一つの特別なケースは、固有値もしくは特異値の和が、X の対角要素の和 $(trace(X))$ になるということである。面白いことに、矩形行列は固有構造を持たないにもかかわらず、その矩形行列 (非対称、非正方) の基本構造が、固有値分解で得られる。矩形行列 X は、二つの対称行列として表現することができ、そして、それらの固有構造が、X の基本構造を得るために用いることができるのである。もし、$XX^T = UDU^T$ で $X^T X = VDV^T$ であるなら、$X = UD^{1/2}V^T$ である。つまり:

> 行列 X の特異値分解は、XT^X の固有値システムによって得られる情報を提供する。より特別なこととして、しかしながら、特異値分解をもちいるほうが好ましい理由がある。第一に、それは、$X^T X$ というようななクロス積行列ではなく、われわれが注目しているデータ行列 X に直接適用されるものであるということである。.... そして.... 与えられた行列についての固有値システムと特異値分解が、数学的に等価であるのだとしても、計算的には、そうではない、ということがある。$X^T X$ に対する、固有値システムよりも数値的に安定した解を導く、X の特異値分解を行うアルゴリズムが存在する。とりわけ、X のランクが、X の次元よりも小さい場合において、である。(Belsey,Kuh and Welsch 1980:99)

2 変換 (Transformations)

主成分分析 (PCA)、多次元尺度法 (MDPREF)、そして対応分析 (CA) のすべては、データ行列における基本構造を見出すことを行う。これらは、同じ分解アルゴリズム、つまり SVD を共有しており、異なっているのは、データに対する処理前[*2]変換と、処理後の潜在ベクトルに対する変換である。明示的であれ暗黙的であれ、これらの方法は、データを分析の途中か、もしくは先立って、データを変換する。例えば、共分散行列や相関行列に対して行われる PCA は、暗黙的に、平均補正、もしくは標準化 (standardization) を、それぞれに行う。MDPRED は、データに対する、平均補正、標準化、そして、二重中心化を先行して行う。CA においては、各値は、対応する周辺度数の幾何平均によって除することになる。これらの変換が各手法おける主要な相違点であるので、この節においては、いくつかの可能な変換を整理しておく。

変換は、行変数もしくは列変数において、または、行と列において、もしくは、行変数間のもしくは列変数間の対になったパターンに対して行われる。おそらく最もお馴染みのものは、正規化

[*2] 訳註：特異値分解

(normalization) は標準化 (standardizaion) であろう。変数を標準化すること、それは、線形変換よる変換であるので、平均がゼロになり標準偏差が 1 になる。定数の加算、減算は、変数の平均にのみ影響するということから、各変数は、新たな平均がゼロになるように変換されるので、平均からの偏差は、以下のようになる。

$$x^*_{ij} = x_{ij} - \bar{X}_i \tag{2.1}$$

定数の乗算、除算は、平均と標準偏差の両方に影響するので、変数 X が、標準偏差の逆数 (1/ σ x) をかけた場合、新たな標準偏差は、1 になる。それゆえ、標準化の線形変換は:

$$x^*_{ij} = \frac{x_{ij} - \bar{X}_i}{\sigma_i} \tag{2.2}$$

変数もまた、空間に投影された時に単位長になるように変換される。この変換は、変数を共通の尺度で再表現するための標準化と似ている。空間的には、単位長ベクトル (刺激点) の終点が、n 次元空間の表面に現れる。そして、2 次元では、それらは弧を描く。単位長に、変数を変換、もしくは正規化 (normalize) するには、

$$x^*_{ij} = \frac{x_{ij}}{(\sum_i x^2_{ij})^{1/2}} \tag{2.3}$$

標準化と単位正規化 (norming) は、定数による違いがあるだけである。観測値の数を (記述的標準偏差の時は n で、推測公式が使われる時は $(n-1)$ で) の平方根で割られて標準化された変数は、単位長ベクトルになる。そして、単位長ベクトルの要素が観測値の数 (n か $n-1$) の平方根をかけられた時は、標準化となる。

カテゴリカルまたは度数データに対する共通の変換は、観測値の総和をもってカテゴリー和を除して求められる。クロス分類データの表において、元のエントリーは、割合 (もしくはパーセント) に変換される。

$$x^*_{ij} = \frac{x_{ij}}{\sum_i x_{ij}} \qquad \text{もしくは} \qquad x^*_{ij} = \frac{x_{ij}}{\sum_j x_{ij}} \tag{2.4}$$

最もシンプルは変換は「リフレクション」である。例えば、1 を最大、k を最小として収集されたするデータは、-1 をかけることによって、もしくは定数 $(k+1)$ を除する変換を施すことができる。この変換は、順序を逆転し、最大数が「もっとも」を最小数「すくない」を指示するようになる。

行と列の双方を変換したい場合もある。これは、同時的もしくは反復的に行われる。行もしくは列の変数は、式 2.1 を用いて、それぞれの平均からの偏差への変換される。行および列両方の平均の除去は、二段階で、もしくは「二重中心化」によって一段階で行われる。二重中心化は、平均効果を取り除き、偏差は変更なしで残す。

$$x^*_{ij} = x_{ij} - \bar{X}_i - \bar{X}_j + \bar{\bar{X}}_{ij} \tag{2.5}$$

標準化 (式 2.2) と割合変換 (式 2.4) の両方は、反復的に遂行できる。データ行列の行と列に対する標準化は、行と列の平均がゼロになり、行内、列内の標準偏差が 1 になるまで反復できる。行と列への割合変換を反復的に行うことは、「iterative propostional fitting」(反復比率適合?) と呼ばれる。これは、「同等」のサイズを反復し、テーブル内の unaffected な連関パターンを残すことによって、行と列における差異を取り除くために使われる (DixonmBrown, et al., 1979:270 Mosteller, 1968)。反復割合適合 (iterative proportional fitting) は、対数線形モデルにおいて、期待値をみつけるために最もよく用いられる (Bishop, Fienberg, and Holland, 1975; Fienberg)。

同じ効果を持つ変換が、周辺度数合計の違いを除去し、すべてのセルを比率で表現する。この「平方根」変換は、各セルを対応する周辺度数合計の幾何平均で割る。

$$x^*_{ij} = \frac{x_{ij}}{(\sum_i x_{ij} \sum_j x_{ij})^{1/2}} = \frac{x_{ij}}{(x_{i.} x_{.j})^{1/2}} \tag{2.6}$$

同じような「パターン」の対としてのデータを再表現する「変換」もある。これらの変換は、通常クロス積の結果である。行列表記において、行列のクロス積は、行列 X かける X の転置となる。XX^T は、行変数に対するクロス積。そして、X^TX は列変数に対するクロス積である。

共分散行列は、平均補正行列 (mean corrected Matrix) のクロス積から得られる。もし、元データが列平均で補正されているなら、クロス積は、次ように得られる。

$$B = X_d^T X_d \tag{2.7}$$

列変数に対する共分散行列は、行の観測値でクロス積を除することで得られる。

$$C = \frac{1}{n} B = \frac{1}{n} X_d^T X_d \tag{2.8}$$

が、X の列変数間の共分散行列を生成する X の行に対する共分散行列は、行平均ち対する最初の補正によって計算出来るので、

$$C = \frac{1}{m} X_d X_d^T \tag{2.9}$$

となり。ここで、m は X の列数である。

共分散 (c_{ij}) の各変数に対する尺度の違いを補正することで、純粋なパターン測値が得られるりピアソンの相関係数 r は、

$$r_{ij} = \frac{c_{ij}}{\sigma_i \sigma_j}$$

もし、X の列が標準化されているなら、補正行列はクロス積の形で計算され、

$$R = \frac{1}{n} X_s^T X_s = \frac{1}{n} Z_c^T Z_c \tag{2.10}$$

となり、ここで、$X_s(Z_c)$ は、列において、標準化された行列であり、n は、行列の観測数である。そして R は、X における全ての m 列のペア間の相関係数を有する行列となる。同様に、X の行間の相関係数 Q は、

$$Q = \frac{1}{m} X_s X_s^T = \frac{1}{m} Z_r Z_r^T \tag{2.11}$$

となり、ここで $X(Z_r)$ は行について標準化された行列を表し、m は列変数の数を表している。

もし、標準化が、標準偏差の推測公式 (分母を $n-1$) でなされていれば、補正要素の式 2.10.2.21 の $1/n$、$1/m$ は、$1/(n-1)$,$1/(m-1)$ になる。平均補正を暗黙に立脚している共分散であれば、相関係数の計算において、標準化は常に暗黙的である。

本章において、データ行列 X の基本構造は、クロス積 ($X^T X$、$X X^T$) によっては影響を受けないことを見てきた。データの正規化 (normalization) と標準化は、しかしながら、基本構造に影響する。基本構造は、変換されたデータ X^* を代表するのであって、元の観測値 X ではないから、因子化に先行して遂行されるこれらの変換は、基本構造に影響する。