特異値分解について

任意の行列 X を、ふたつのユニタリ行列 U, V とひとつの対角行列 D の内積で表すことができる。これを特異値分解という。たとえば R では svd()という関数で、あっという間に答えをだしてくれる（このとき X には欠損値があってはいけない)。

$$X = UDV^*$$

V^* は V の随伴行列、まあデータが実数なら転置行列 V^T に等しい。この３つの行列って具体的にどんなもので、どうやって（もし自分で計算するのなら）算出できるのかを以下に説明したい。備忘録みたいなもんです、よしなに。

想像するのが簡単なので、X をとりあえず「測定が２項目、サンプルが３検体のデータ」だとする。

$$X = \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{31} & x_{32} \end{pmatrix}$$

これの逆行列とこれの内積をとると、2×2 の対称行列ができる。

$$A = X^*X = \begin{pmatrix} x_{11} & \cdots & x_{31} \\ x_{12} & \cdots & x_{32} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} \\ \vdots & \vdots \\ x_{31} & x_{32} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}^2 + x_{21}^2 + x_{31}^2 & x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} \\ x_{11}x_{12} + x_{21}x_{22} + x_{31}x_{32} & x_{12}^2 + x_{22}^2 + x_{32}^2 \end{pmatrix}$$

この値は何か？　もし X が、それぞれの測定をセンタリングしてあれば、これは共分散の（３検体―自由度１）倍の値になる。さらに測定がスケーリングしてあれば、相関の（３－１）倍になる。測定と測定のあいだの関連性をあらわす値であると見ることができるだろう。

固有値分解について

対称行列 A　を変換する縦ベクトルvについて、こういうベクトルを考える。

$$Av = \lambda v$$

ここでλ はスカラーである。A と内積をとることで、そのベクトルv のスカラー倍にできるベクトルである。

$$(A - \lambda I)v = 0$$

この式で、$v = 0$ 以外のベクトルを探すことにする。すると

$$\det(\mathrm{A}-\lambda\mathrm{I}) = |\mathrm{A}-\lambda\mathrm{I}| = 0$$

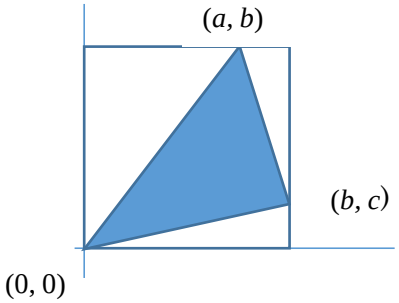
ここで det とはなにか？

$$\mathrm{B} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

と書くとき、

$$\det(\mathrm{B}) = |\mathrm{B}| = ac - b^2$$

である。これは点 (a, c) と原点を対角にもつ長方形と、(b, b) と原点を対角にもつ正方形の面積の差になる。これがゼロになるとき、点 (a, b) と点 (b, c) をむすぶ直線は原点を通る（あるいは、これらの２点と原点でつくる三角形の面積がゼロになる）。



この青い三角形の面積は、白い正方形から３つの三角形を引いたものである。計算するとそれは $(b^2\text{-}ac)/2$ なので、det(B)=0 ならこの三角形には面積がないことになる。

三次元

$$\mathrm{X} = \begin{pmatrix} x_{11} & \cdots & x_{13} \\ \vdots & \ddots & \vdots \\ x_{41} & \cdots & x_{43} \end{pmatrix}$$

$$\mathrm{A} = \mathrm{X}^*\mathrm{X} = \begin{pmatrix} x_{11} & \cdots & x_{41} \\ \vdots & \ddots & \vdots \\ x_{13} & \cdots & x_{43} \end{pmatrix}\begin{pmatrix} x_{11} & \cdots & x_{13} \\ \vdots & \ddots & \vdots \\ x_{41} & \cdots & x_{43} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}^2 + x_{21}^2 + x_{31}^2 + x_{41}^2 & \cdots & x_{11}x_{13} + x_{21}x_{23} + x_{31}x_{33} + x_{41}x_{43} \\ \vdots & \ddots & \vdots \\ x_{11}x_{13} + x_{21}x_{23} + x_{31}x_{33} + x_{41}x_{43} & \cdots & x_{13}^2 + x_{23}^2 + x_{33}^2 + x_{43}^2 \end{pmatrix}$$

面積がないというのはどういうこと？

もうひとつ次元をあげてみる。

$$\mathrm{B} = \begin{pmatrix} a & b & c \\ b & d & e \\ c & e & f \end{pmatrix}$$

と書くとき、

$$\det(\mathrm{B}) = |\mathrm{B}| = \mathrm{a}\begin{vmatrix} d & e \\ e & f \end{vmatrix} - \mathrm{b}\begin{vmatrix} b & c \\ e & f \end{vmatrix} + \mathrm{c}\begin{vmatrix} b & c \\ d & e \end{vmatrix}$$

また

$$\begin{vmatrix} d & e \\ e & f \end{vmatrix} = df - e^2$$

である。これは、各点(a, b, c)、(b, d, e)、(c, e, f) へと原点からひいた３本の線分を３辺にもつ平行６面体の体積である。
そこで、$\det(\mathrm{A} - \lambda\mathrm{I}) = 0$にするためには、

$$\mathrm{A} - \lambda\mathrm{I} = \begin{pmatrix} a-\lambda & b & c \\ b & d-\lambda & e \\ c & e & f-\lambda \end{pmatrix}$$

にある各点$(a\text{-}\lambda, b, c)$、$(b, d\text{-}\lambda, e)$、$(c, e, f\text{-}\lambda)$ が指定する平面が、原点をふくむようになる λ を求めればいいことになる。これはλの三次式になるので、この値が３つ存在する。ちなみにそれぞれのλを引いたAの内積の行列式はゼロになる。三次関数の解の公式はめんどくさいのでパス。　　これきっと任意の次元で成立して、定理になってたりするんじゃなかろうか。R でランダム数をつかって計算すると 10 次元くらいまでは det がゼロになるが、あまり大きくするとそうでもなくなる、私はそれは計算精度のせいだと思っているんだけど違うかもしれない。解の公式があるのは４次までらしいから、あとで書くような泥臭い方法ではｎ次のことは証明できない。

平面のとき直線になって面積がなくなり、立体のときに平面になって体積がなくなる。このとき、その直線なり平面の上で、データの原点からの距離が、もっとも効率よく表されることになる。λ の値は、$\mathrm{A}v = \lambda v$ から、v によって変換した A の大きさであることが直感的にわかる。

またλ が引かれる因子は（Aが対称行列なので）、それぞれ、因子の二乗和になるところである（たとえばAが相関のマトリクスであるなら、すべて１になるところである）。

固有値分解について（続き）

さて、二次元に話をもどそう。$\det(\mathrm{A}-\lambda\mathrm{I})=0$にするためには、

$$\mathrm{A}-\lambda\mathrm{I}=\begin{pmatrix} x_{11}^2+x_{21}^2+x_{31}^2-\lambda & x_{11}x_{12}+x_{21}x_{22}+x_{31}x_{32} \\ x_{11}x_{12}+x_{21}x_{22}+x_{31}x_{32} & x_{12}^2+x_{22}^2+x_{32}^2-\lambda \end{pmatrix}$$

が指定する２つの点と原点が直線上にのるようにすればいい。これはλ についての二次式になる。簡単に

$$\mathrm{A}-\lambda\mathrm{I}=\begin{pmatrix} a-\lambda & b \\ b & c-\lambda \end{pmatrix}=0$$

とかくと、二次式の解の公式から$\lambda=\frac{1}{2}\left(a+c\pm\sqrt{(a+c)^2-4ac+4b^2}\right)$　が導かれる。

λのそれぞれの値をλ_1 とλ_2 と書くと、これらをそれぞれ引いた行列　$\mathrm{A}-\lambda_1$と　$\mathrm{A}-\lambda_2$ の内積はゼロ行列になる。

$$(\mathrm{A}-\lambda_1\mathrm{I})^{\mathrm{T}}(\mathrm{A}-\lambda_2\mathrm{I})=\begin{pmatrix} a-\lambda_1 & b \\ b & c-\lambda_1 \end{pmatrix}\begin{pmatrix} a-\lambda_2 & b \\ b & c-\lambda_2 \end{pmatrix}=\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$$

このうちαは

$$\alpha=a^2-(a+c)a+a^2+ac-b^2+b^2=0$$

同様に

$$\beta=b(a+c-(a+c))=0$$

$$\gamma=b^2-(a+c)a+c^2+ac-b^2=0$$

そこで

$(\mathrm{A}-\lambda_1\mathrm{I})^{\mathrm{T}}(\mathrm{A}-\lambda_2\mathrm{I})=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$　となる。

当然、その行列式　$\det(\,(\mathrm{A}-\lambda_1\mathrm{I})(\mathrm{A}-\lambda_2\mathrm{I}\,)\,)=0$ になる。転置行列の内積がゼロになるので、$\mathrm{A}-\lambda_1\mathrm{I}$と$\mathrm{A}-\lambda_2\mathrm{I}$ は独立で、これらの行列が指定する直線は直行する。

必然的に、それぞれを回転させるためのv もまた直行することになる。

このλ を Aの eigenvalue ないし固有値という。またそれぞれのλについてv を求めることができる。このv はかならずしも一意に求まらない。しかし、正規化する：すなわちvの各要素の二乗和を１とすることで値がきまる。
このベクトルのことを A の eigenvector ないし固有ベクトルという。

さて A と v の内積はひとつのベクトルになるが、これは$\mathrm{A}v=\lambda v$ つまり、方向が v で長さがλ であるベクトルになる。

三次式だと

$\mathrm{A}v = \begin{pmatrix} av_\mathrm{x} & bv_\mathrm{y} & cv_\mathrm{z} \\ bv_\mathrm{x} & dv_\mathrm{y} & ev_\mathrm{z} \\ cv_\mathrm{x} & ev_\mathrm{y} & fv_\mathrm{z} \end{pmatrix}$ について、

$(\mathrm{A}-\lambda\mathrm{I})v = \begin{pmatrix} (a-\lambda)v_\mathrm{x} & bv_\mathrm{y} & cv_\mathrm{z} \\ bv_\mathrm{x} & (d-\lambda)v_\mathrm{y} & ev_\mathrm{z} \\ cv_\mathrm{x} & ev_\mathrm{y} & (f-\lambda)v_\mathrm{z} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ だったので

$\mathrm{A}v = \mathrm{A}v - (\mathrm{A}-\lambda\mathrm{I})v = \lambda \begin{pmatrix} v_\mathrm{x} \\ v_\mathrm{y} \\ v_\mathrm{z} \end{pmatrix}$ である。

λ を大きい順にならべて $\lambda_{1,2,3}$ として、それぞれに対応する v を $v_{1,2,3}$ とするとき、
$\mathrm{V} = (v_1 \quad v_2 \quad v_3)$ という行列を考える。これと A の内積をとると、
AV の各列は、それぞれ $v_{1,2,3}$ によって変換したベクトルになっていて、その長さは $\lambda_{1,2,3}$ である。

そこで $\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$ という対角行列を考えると

$$\mathrm{AV} = \mathrm{V}\Lambda$$

となる。

さて V にはおもしろい性質がある。まず $\mathrm{V}^*\mathrm{V} = \mathrm{I}$ である、これは $v_{1,2,3}$ が正規化されていること、そしてそれぞれが独立であることから導かれる。そこで $\mathrm{V}\mathrm{V}^*\mathrm{V} = \mathrm{V}$ なので、$\mathrm{V}\mathrm{V}^*\mathrm{V}\mathrm{V}^{-1} = \mathrm{V}\mathrm{V}^{-1} = \mathrm{I}$ だから $\mathrm{V}\mathrm{V}^* = \mathrm{I}$ となる。つまり V はユニタリ行列であり、V の各行もまた正規化されていることになる。
そこで、

$$\mathrm{A} = \mathrm{A}\mathrm{V}\mathrm{V}^* = \mathrm{V}\Lambda\mathrm{V}^*$$

これを A の固有値分解という。

ふたたび特異値分解

$$\mathrm{X} = \mathrm{U}\mathrm{D}\mathrm{V}^*$$

この V は、

$$\mathrm{A} = \mathrm{X}^*\mathrm{X} = \mathrm{V}\Lambda\mathrm{V}^*$$

これと同じものだ。（ベクトルなので、符号が逆に算出されることはあるにせよ）。

同様に、

$$XX^* = U\Lambda U^*$$

でもある。そして

$$D = \sqrt{\Lambda}$$

である。

PCA　主成分分析について

$$X = UDV^*$$

$PC_{sample} = XV = UD$ である。これはどういうものか？

V がユニタリ行列なので、これは X を原点中心で回転させたものと考えることができる。U には、それぞれの軸において、それぞれのサンプルがどの位置にあるのかが指定されている。これは正規化されているので、その位置は、二乗値総和が 1 になるようになっている。サンプルのお互いの位置関係はここに書かれている。その縮尺が D に書かれている。そこで、この式において、V は回転を（そして新しい軸の方向を）、U は各サンプルの相対的な位置を、D は縮尺を（分散でなくて標準偏差で）記録している。

同様に、$PC_{item} = X'U = VD$ である。これは？

X の転置行列を U で回転させたものである。X を、サンプル数だけの時限をもった、測定項目のデータだと目して理解するとこの形になる。V には、それぞれの軸における項目の位置が書かれ、D はその縮尺である。こんどは U が軸の方向を指定する。

バイプロットについて

ふたつの主成分 PC_{sample} と PC_{item} を同一平面に記述できると、なにかと便利である。これを biplot というのだけど、これは本来的な定義(?) では UD と V を、適当な縮尺を与えながら作図することになっている（Jackson, 1991 とか）。

せっかく V にも D を与えたのだから、それをそのまま表せたほうが都合はいい。そのためには、しかし、それぞれの大きさの違いが問題になる。

X を m 行 n 列の行列だとすると、

$X=UDV^*$ とするとき、U は m 行 m 列の、V は n 行 n 列のユニタリ行列になる。D の値が共通なので、PC_{sample} と PC_{item} は、そのスケールが異なっている；ユニタリ行列は縦横ともに二乗和が 1 になるからだ。

サンプルをアイテム数の平方根で、アイテムをサンプル数の平方根で除することで、スケールを揃えることができる。

$PC_{sample} = XV = UD$

にある、いずれかの PC のサンプルの値は、n 列の X と n 列の V の積の和である。大きさを表す D は m サンプルぶんの標準偏差のかたちになっているので、これを n のアイテム数の平方根で除したものと、同様に PC_{item} を m の平方根で除したものは、データのなかのちらばりを nm の平方根で除した、いわば平均的なちらばりをもつことになる。

$sPC_{sample} = UDn^{1/2}$

$sPC_{item} = VDm^{1/2}$