主成分分析と行列の特異値分解

Ver.2022.09.04

松浦 真也

主成分分析は，データの相関係数行列の固有値，固有ベクトルを用いて解釈可能であった．この資料では，少し別の言葉（行列の特異値分解）を用いて，主成分分析に対する理解を深めたい．以下の各ステップに従い，各自で考察してみよう．

ステップ 0 まず，準備として，以下の設定を行う．

- A を $m \times n$ 実行列（実数を成分にもつ行列）とする．
- 簡単のため，$m < n$ とする（この仮定は本質的ではない）．
- 以下の各ステップでは，それより前のステップで定義された記号は，そのまま引き継いでいるものとする．例えば，ステップ 1 で定義された記号は，ステップ 2 以降でも，引き続き有効である．

ステップ 1 $P^{-1}A\,{}^tAP$ が実対角行列となるような m 次直交行列 P が存在する．その理由を考えてみよう．ただし，${}^tA$ は A の転置行列である（一般に，^{t} は行列やベクトルの転置を表すものとする．）．

ステップ 2 ${}^tAP$ の列ベクトルを $\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}$ とする．つまり，

$$ {}^tAP = \Big[\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}\Big] $$

とする．このとき，各 $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$ に対し，${}^t\boldsymbol{v}_i\boldsymbol{v}_j$ の値を求めてみよう．必要があれば，$P^{-1}A\,{}^tAP$ の対角成分 $d_1, d_2, \ldots, d_m$ を用いて表してよい．

ステップ 3 任意の i $(i = 1, 2, \ldots, m)$ に対して，$d_i \geq 0$ であることを示してみよう．さらに，$d_i = 0$ なら $\boldsymbol{v}_i = 0$ であることも示してみよう．

ステップ 4 n 次元空間 $\mathbb{R}^n$ の正規直交基底 $\boldsymbol{q_1}, \ldots, \boldsymbol{q_n}$ で，次を満たすものが存在することを示してみよう．

$$ i = 1, 2, \ldots, m \text{ に対して}, \qquad d_i \neq 0 \qquad \Rightarrow \qquad \boldsymbol{q_i} = \frac{1}{\sqrt{d_i}}\boldsymbol{v_i}. $$

m と n の区別に注意すること．また，$m < n$ としていることにも注意．

ステップ 5 $\boldsymbol{q_1}, \ldots, \boldsymbol{q_n}$ を列ベクトルとする n 次正方行列を Q とする．つまり，

$$ Q = \Big[\boldsymbol{q_1}, \boldsymbol{q_2}, \ldots, \boldsymbol{q_n}\Big] $$

とする．このとき，

$$P^{-1}AQ = \begin{bmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_m} \end{bmatrix}$$

であることを示してみよう．ただし，行列の空欄になっている成分は，すべて 0 とする．

ステップ 6 ここまでの話をまとめると，任意の $m \times n$ 実行列 A は，

$$A = P \begin{bmatrix} \sqrt{d_1} & & & \\ & \sqrt{d_2} & & \\ & & \ddots & \\ & & & \sqrt{d_m} \end{bmatrix} Q^{-1}$$

と分解できる．ここで，P は m 次直交行列，Q は n 次直交行列，$d_1, d_2, \ldots, d_m$ は非負の実数である．この分解を，行列 A の特異値（とくいち）分解と呼ぶ．行列の対角化は対称行列など，特殊な行列に対してのみ実行可能なのに対し，特異値分解は，正方行列に限らず，任意の行列に対して実行可能である．

以上を踏まえ，主成分分析について考察したい．規格化されたデータ z_{ij} を (i, j) 成分にもつ行列を Z とするとき，主成分分析と Z の特異値分解との間に，どのような関係があるか考察してみよう．

ヒント：A の特異値分解に現れる $d_1, d_2, \ldots, d_m$ や P は，$A\,{}^tA$ の何に対応するか？