

## はじめに

主成分分析 (Principal Component Analysis: PCA) を幾何学的に理解する。幾何学的理解のためには、2次元で考えるとわかりやすい。

目標:  $n$  個のデータの  $(x, y)$  座標が次の行列で与えられたとき、データを単位ベクトル  $a = \begin{pmatrix} a_x \\ a_y \end{pmatrix}$  に射影した時の分散が最も大きくなる  $a$  (第1モード) をみつける。ただし  $|a| = 1$  であり、データ  $X$  は事前に  $(x, y)$  のどちらとも平均で減算されている (平均が 0 である) ものとする。

$$X = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix}$$

## ベクトルの射影

上記のデータを単位ベクトル  $a$  に射影した時の長さを考える。例えば  $a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  であれば、データをベクトル  $a$  に射影した長さは、それぞれの  $x$  座標となる。この操作はベクトル同士の内積によって説明される。あるデータ点  $X_i = (x_i, y_i)$  とベクトル  $a$  の内積は

$$a^T X_i = (a_x, a_y) \begin{pmatrix} x_i \\ y_i \end{pmatrix} = a_x x_i + a_y y_i$$

となる。別の表現をするならば、ベクトル  $a$  と  $X_i$  のなす角を  $\theta$  とすると、

$$a^T X_i = (a_x, a_y) \begin{pmatrix} x_i \\ y_i \end{pmatrix} = |a^T| |X_i| \cos \theta = |X_i| \cos \theta$$

であり、これが単位ベクトル  $a$  に対する射影であることも幾何学的に理解できる。

この操作を全データに対して同時に行い、 $n$  個のデータを単位ベクトル  $a$  に射影した長さは

$$a^T X = (a_x, a_y) \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix} = (a_x x_1 + a_y y_1, \dots, a_x x_n + a_y y_n)$$

によって与えられる。ベクトル  $a^T$  の形状 (行数, 列数) は  $(1, 2)$  であり、 $X$  の形状は  $(2, n)$  であるから、 $a^T X$  の形状は  $(1, n)$  であり、ベクトル  $a$  に射影した長さが  $n$  個のデータそれぞれで計算された結果をまとめたベクトルになっている。

## 射影後の分散

データ  $X$  をベクトル  $a$  に射影した時の長さが  $a^T X$  で与えられた。ここで目的を言い換えると「 $a^T X$  の分散が最も大きくなる  $a$  をみつける」こととなる。 $a^T X$  の分散  $V$  は次の式で与えられる:

$$V \equiv \frac{1}{n-1} (a^T X)^2 = \frac{1}{n-1} (a^T X) (a^T X)^T = \frac{1}{n-1} a^T X X^T a = a^T S a$$

ここで、 $S \equiv \frac{1}{n-1} X X^T$  である。なお、 $V$  はスカラー値となる。

## 共分散行列の対角化

ここで、分散  $V$  の計算に出現した  $S \equiv \frac{1}{n-1}XX^T$  について考える。これは一般に「分散共分散行列」または単に「共分散行列」と呼ばれる対称行列である。 $S$  を成分で書き下してみると、

$$S = \frac{1}{n-1}XX^T = \frac{1}{n-1} \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{pmatrix} \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix} = \frac{1}{n-1} \begin{pmatrix} \sum_{i=1}^n x_i x_i & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i x_i & \sum_{i=1}^n y_i y_i \end{pmatrix}$$

となり、次のことがわかる：

1.  $S$  は2次の対称行列 ( $S^T = S$  が成り立つ行列) である
2.  $S$  は半正定値行列であるため、その固有値は全て非負である<sup>1</sup>
3.  $S$  の対角成分は  $x$  の分散と  $y$  の分散である ( $X$  の平均は 0 なので、 $x_i$  と  $y_i$  は平均からの偏差となっているため)
4.  $S$  の非対角成分は  $x$  と  $y$  の共分散である

<sup>1</sup>共分散行列が半正定値行列であることは別途証明を要するが、PCAを理解するには  $S$  の「固有値が全て非負になること」を認めたことにして読み進めても良い。

上記の 3, 4 番目の事実が、この行列が共分散行列と呼ばれる所以である。また、元データの次元数を  $d$  とすると、 $S$  はデータ数  $n$  に関係なく  $d \times d$  の対称行列となる (ここでは  $d = 2$ )。

$S$  は対称行列なので、正規直交行列  $P$  と対角行列  $\Lambda$  を用いて固有値分解することができる：

$$S = P\Lambda P^T$$

ここで、 $P \equiv (v_1, v_2)$ 、 $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$  であり、 $\lambda_i$  と  $v_i$  はそれぞれ  $i$  番目に大きい固有値と固有ベクトルである。 $S$  は対称行列なので  $P$  は正規直交行列である、つまり  $v_1$  と  $v_2$  は互いに直交し、そのノルムは 1 である (重要)。

これを用いると、 $V$  は

$$V = a^T S a = a^T P \Lambda P^T a$$

と書き換えられる。さらに  $u \equiv \begin{pmatrix} u_x \\ u_y \end{pmatrix} = P^T a$  とおくと、

$$V = u^T \Lambda u = (u_x, u_y) \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix} = \lambda_1 u_x^2 + \lambda_2 u_y^2$$

となる。ここで、 $u$  の各成分は

$$\begin{pmatrix} u_x \\ u_y \end{pmatrix} = P^T a = \begin{pmatrix} v_1^T \\ v_2^T \end{pmatrix} a = \begin{pmatrix} v_1^T a \\ v_2^T a \end{pmatrix}$$

であるから、結局  $V$  は

$$V = \lambda_1 (v_1^T a)^2 + \lambda_2 (v_2^T a)^2$$

となる。

## 分散が最大となる方向

ここまで、任意の単位ベクトル  $a$  に対して射影した長さの分散  $V$  を変形してきた。これを最大化する  $a$  を知ることが目的である。ここで、 $V$  に現れる  $v_i^T a$  の形をよくみると、これはベクトル  $a$  をベクトル  $v_i$  に射影した長さ、つまり  $S$  の固有ベクトル  $v_i$  と  $a$  の内積を表していることに気付く（上述の  $a^T X_i$  と同じ形）。

よって、単位ベクトル  $a$  と固有ベクトル  $v_1, v_2$  のなす角をそれぞれ  $\theta_1, \theta_2$  と表すと、分散  $V$  は

$$V = \lambda_1 \cos^2 \theta_1 + \lambda_2 \cos^2 \theta_2$$

と書くことができる。なお、 $0^\circ \leq \theta_1, \theta_2 < 180^\circ$  を満たす。

$v_1$  と  $v_2$  は直交するので、 $v_1, v_2$  座標空間における第一象限に  $a$  がある場合、 $\theta_1 + \theta_2 = 90^\circ$  と表すことができ、 $\theta_2$  を消去すると

$$V = \lambda_1 \cos^2 \theta_1 + \lambda_2 \cos^2(90^\circ - \theta_1) = (\lambda_1 - \lambda_2) \cos^2 \theta_1 + \lambda_2$$

となる。このとき、 $(\lambda_1 - \lambda_2) \geq 0$  であるから、 $V$  が最大となる  $\theta_1$  は  $0$  であり、これは単位ベクトル  $a$  が固有ベクトル  $v_1$  の方向を向くときに分散が最大となることを意味する。なお、 $a$  が  $v_1, v_2$  座標空間における第一象限ではない象限にあった場合でも同様の結論を得ることができる。

以上より、任意のデータに対して分散を最大化させる方向は、そのデータの共分散行列の最も大きい固有値に対応する固有ベクトルの方向であることが示された。