作成日：2019 年 9 月 7 日
慶應義塾大学理工学部物理学科
岡崎健人

主成分分析　自分用教科書

1　２次元の場合

n 個の 2 次元データ $\boldsymbol{x}_i = [x_i, y_i]^\mathsf{T}$ $(i = 1, 2, \ldots, n)$ がある。これを $z = ax + by$ のように変換して、z 軸上でデータの分散が最大になるように係数 $\boldsymbol{a} = [a, b]^\mathsf{T}$ を決めたい。z 軸（この軸を第 1 主成分という）に射影したデータを $z_i = ax_i + by_i = \boldsymbol{a}^\mathsf{T}\boldsymbol{x}_i$ とする。z_i の平均 $\overline{z}$ は

$$\overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{a}^\mathsf{T}\boldsymbol{x}_i = \boldsymbol{a}^\mathsf{T}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = \boldsymbol{a}^\mathsf{T}\overline{\boldsymbol{x}}$$

となる。ここに $\overline{\boldsymbol{x}} = [\overline{x}, \overline{y}]^\mathsf{T}$ である。z の分散は

$$\begin{aligned}
s_z^2 &= \frac{1}{n}\sum_{i=1}^{n}(z_i - \overline{z})^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{a}^\mathsf{T}\boldsymbol{x}_i - \boldsymbol{a}^\mathsf{T}\overline{\boldsymbol{x}}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[a(x_i - \overline{x}) + b(y_i - \overline{y})\right]^2 \\
&= a^2\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 + 2ab\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) + b^2\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 \\
&= a^2 s_x^2 + 2ab s_{xy} + b^2 s_y^2 \\
&= [a, b]\begin{bmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} \\
&= \boldsymbol{a}^\mathsf{T} S \boldsymbol{a}
\end{aligned}$$

となる。S は標本分散共分散行列と呼ばれる。よって問題は $s_z^2 = \boldsymbol{a}^\mathsf{T}S\boldsymbol{a}$ が最大値を取る $\boldsymbol{a}$ を求めることである。ただしこのままだと $\|\boldsymbol{a}\| \to \infty$ となり s_z^2 が発散してしまうので、$\boldsymbol{a}^\mathsf{T}\boldsymbol{a} = 1$ という条件をつける。この制約条件下での最大化問題はラグランジュの未定乗数法により解くことができる。未定乗数を λ として

$$F(\boldsymbol{a}, \lambda) = \boldsymbol{a}^\mathsf{T}S\boldsymbol{a} + \lambda\left(1 - \boldsymbol{a}^\mathsf{T}\boldsymbol{a}\right)$$

とおき、これを $\boldsymbol{a}$ で微分したものがゼロベクトルとなるような $\boldsymbol{a}$ を求めるのである。$\boldsymbol{\nabla}_{\boldsymbol{a}} = [\partial_a, \partial_b]^\mathsf{T}$ とすると $\boldsymbol{\nabla}_{\boldsymbol{a}}F(\boldsymbol{a}, \lambda) = 2S\boldsymbol{a} - 2\lambda\boldsymbol{a} = \boldsymbol{0}$ より

$$S\boldsymbol{a} = \lambda\boldsymbol{a}.$$

ただし $\boldsymbol{\nabla}_{\boldsymbol{a}}\left(\boldsymbol{a}^\mathsf{T}S\boldsymbol{a}\right) = 2S\boldsymbol{a},\ \boldsymbol{\nabla}_{\boldsymbol{a}}\left(\boldsymbol{a}^\mathsf{T}\boldsymbol{a}\right) = 2\boldsymbol{a}$ となることを用いた $\boldsymbol{a}$ は標本分散共分散行列 S の固有ベクトルであったことがわかる。また左から $\boldsymbol{a}^\mathsf{T}$ をかけて $\boldsymbol{a}^\mathsf{T}\boldsymbol{a} = 1$ を使えば $s_z^2 = \boldsymbol{a}^\mathsf{T}S\boldsymbol{a} = \lambda,$ ゆえに最大化したかった分散は固有値に一致することがわかる。最大となる固有値を $\lambda_1,$ それに属する固有ベクトルを $\boldsymbol{a}_1$ とすれば第 1 主成分は $z_1 = {\boldsymbol{a}_1}^\mathsf{T}\boldsymbol{x}$ となる。

次に第 1 主成分と直交し、その分散が最大となるような軸 w（第 2 主成分）を求める。2 次元の場合は直交するという条件のみから決められるが、より多次元のデータを主成分分析すると直交するという条件のみからは決められない。その係数ベクトルを $\boldsymbol{b}$ とすると、その大きさが 1 であり（$\boldsymbol{b}^\mathsf{T}\boldsymbol{b} = 1$）、$\boldsymbol{a}_1$ と直交し

($\boldsymbol{a}_1{}^\top \boldsymbol{b} = 0$)、$w$ 方向の分散 $s_w^2 = \boldsymbol{b}^\top S \boldsymbol{b}$ を最大化する。ふたたびラグランジュの未定乗数法を用いる。λ, γ を未定乗数として、

$$F(\boldsymbol{b}, \lambda, \gamma) = \boldsymbol{b}^\top S \boldsymbol{b} + \lambda\Big(1 - \boldsymbol{b}^\top \boldsymbol{b}\Big) + \gamma \boldsymbol{a}_1{}^\top \boldsymbol{b}$$

という関数を微分したものをゼロとおくと、$\boldsymbol{\nabla}_{\boldsymbol{b}} F(\boldsymbol{b}, \lambda, \gamma) = 2S\boldsymbol{b} - 2\lambda\boldsymbol{b} + \gamma\boldsymbol{a}_1 = \boldsymbol{0}$. これに左から $\boldsymbol{a}_1{}^\top$ をかけ、$S\boldsymbol{a}_1 = \lambda_1 \boldsymbol{a}_1$ より $\boldsymbol{a}_1{}^\top S = \boldsymbol{a}_1{}^\top \lambda_1$, $\boldsymbol{a}_1{}^\top \boldsymbol{b} = 0$, $\boldsymbol{a}_1{}^\top \boldsymbol{a}_1 = 1$ となることを用いれば $\gamma = 0$ を得る。したがって $\boldsymbol{b}$ は再び S の固有ベクトルであることがわかる。その固有値は 2 番目の固有値 λ_2, 固有ベクトルは $\boldsymbol{a}_2$, w の分散は $s_w^2 = \boldsymbol{a}_2^\top S \boldsymbol{a}_2 = \lambda_2$ となる。

2　一般次元の場合

p 次元の変数 $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]$ について、n 個のデータ $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ がある。平均のベクトルを $\overline{\boldsymbol{x}}$ とする。また j 成分の分散を s_j^2, j 成分と k 成分の共分散を s_{jk} と書く。また l 番目に分散を大きくするようにとった軸（第 l 主成分）を $y^{(l)} = \boldsymbol{a}^{(l)}{}^\top \boldsymbol{x}$ と書くことにする。i 番目のデータ $\boldsymbol{x}_i$ を変換すると $y_i^{(l)} = \boldsymbol{a}^{(l)}{}^\top \boldsymbol{x}_i$ となる。係数ベクトルは $\boldsymbol{a}^{(l)} = \left[a_1^{(l)}, a_2^{(l)}, \ldots, a_p^{(l)}\right]^\top$ である。$y^{(l)}$ の分散は

$$s_{y^{(l)}}^2 = \frac{1}{n} \sum_{i=}^{n} \left(\boldsymbol{a}^{(l)}{}^\top \boldsymbol{x}_i - \boldsymbol{a}^{(l)}{}^\top \overline{\boldsymbol{x}}\right)^2$$

証明. n 次元のベクトル $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^\mathsf{T}$ と n 次対称行列 A について、$\boldsymbol{\nabla} = [\partial_1, \partial_2, \ldots, \partial_n]^\mathsf{T}$ とすると $\boldsymbol{\nabla}\left(\boldsymbol{x}^\mathsf{T} A \boldsymbol{x}\right) = 2A\boldsymbol{x}$, $\boldsymbol{\nabla}\left(\boldsymbol{x}^\mathsf{T}\boldsymbol{x}\right) = 2\boldsymbol{x}$ となることを示す。まず $\boldsymbol{\nabla}\left(\boldsymbol{x}^\mathsf{T}\boldsymbol{x}\right) = 2\boldsymbol{x}$ のほうが簡単そうなのでこちらから証明する。アインシュタインの縮約を用いれば $\boldsymbol{x}^\mathsf{T}\boldsymbol{x} = x_i x_i$ なので

$$\boldsymbol{\nabla}\left(\boldsymbol{x}^\mathsf{T}\boldsymbol{x}\right) = \partial_j (x_i x_i)\boldsymbol{e}_j = 2x_i(\partial_j x_i)\boldsymbol{e}_j = 2x_i \delta_{ij}\boldsymbol{e}_j = 2x_i \boldsymbol{e}_i = 2\boldsymbol{x}.$$

また $A\boldsymbol{x} = (a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{jn}x_n)\boldsymbol{e}_j = a_{ji}x_i\boldsymbol{e}_j$ であるから $\boldsymbol{x}^\mathsf{T} A\boldsymbol{x} = x_j a_{ji} x_i = a_{ij}x_i x_j$. ゆえに

$$\begin{aligned}
\boldsymbol{\nabla}\left(\boldsymbol{x}^\mathsf{T} A\boldsymbol{x}\right) &= \partial_k (a_{ij}x_i x_j)\boldsymbol{e}_k = a_{ij}\partial_k (x_i x_j)\boldsymbol{e}_k = a_{ij}x_i\delta_{kj}\boldsymbol{e}_k + a_{ij}\delta_{ik}x_j\boldsymbol{e}_k \\
&= a_{ij}x_i\boldsymbol{e}_j + a_{ij}x_j\boldsymbol{e}_i = a^\mathsf{T}_{ji}x_i\boldsymbol{e}_j + a_{ij}x_j\boldsymbol{e}_i \\
&= A^\mathsf{T}\boldsymbol{x} + A\boldsymbol{x} = \left(A^\mathsf{T} + A\right)\boldsymbol{x} \\
&= 2A\boldsymbol{x}.
\end{aligned}$$

ただし A が対称行列であること $A^\mathsf{T} = A$ を用いた。 □