主成分分析演習 (1)

2009 年 10 月 7 日 古川徹生

1　ベクトルを用いた楕円の方程式と二次形式

1.1　単位円の方程式

単位円（半径 1 の円）の方程式 ${x_1}^2 + {x_2}^2 = 1$ を考える．ベクトル $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ を使えば，この方程式は

$$\begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathbf{x}^T \mathbf{x} = 1 \tag{1}$$

と書くことができる．なお単位行列 $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ を使うと (1) は

$$\mathbf{x}^T I \mathbf{x} = 1 \tag{2}$$

となる．I は対称行列であることに注意．このように実対称行列の両側からベクトル $\mathbf{x}^T, \mathbf{x}$ をかけるとスカラーが得られる．これを**二次形式**という．

1.2　任意サイズの円の方程式

次に半径 a の円の方程式 ${y_1}^2 + {y_2}^2 = a^2$ を考えてみる．二次形式を使えば

$$\mathbf{y}^T I \mathbf{y} = a^2 \tag{3}$$

と書ける．また辺々を a^2 で割れば

$$\frac{y_1^2}{a^2} + \frac{y_2^2}{a^2} = 1 \tag{4}$$

であるから，

$$\mathbf{y}^T \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{a^2} \end{pmatrix} \mathbf{y} = 1 \tag{5}$$

と書くこともできる．

1.3　単位円から半径 r 円への拡大写像

半径 1 の円を半径 a に拡大する写像を考えてみよう．この写像は行列を使って次のように表現できる．

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{6}$$

ここで $A = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ とおけば

$$\mathbf{y} = A\mathbf{x} \tag{7}$$

と表される．また逆変換は

$$A^{-1} = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{a} \end{pmatrix} \tag{8}$$

$$\mathbf{x} = A^{-1}\mathbf{y} \tag{9}$$

である．さて，$\mathbf{x}$ は単位円上の点であるから，(1) を満たす．(1) に (9) を代入すると，

$$\mathbf{x}^T\mathbf{x} = 1 \tag{10}$$

$$(A^{-1}\mathbf{y})^T(A^{-1}\mathbf{y}) = 1 \tag{11}$$

$$\mathbf{y}^T A^{-1^T} A^{-1}\mathbf{y} = 1 \tag{12}$$

$$\mathbf{y}^T \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{a^2} \end{pmatrix} \mathbf{y} = 1 \tag{13}$$

となり，(5) と同じ結果が得られる．ここで対角行列 Λ を次のように定義すれば，

$$\Lambda = A^2 = \begin{pmatrix} a^2 & 0 \\ 0 & a^2 \end{pmatrix} \tag{14}$$

$A = \Lambda^{1/2}, A^{-1} = \Lambda^{-1/2}$ であり，(13) は

$$\mathbf{y}^T \Lambda^{-1} \mathbf{y} = 1 \tag{15}$$

と書くことができる．

1.4　単位円から楕円への写像 (1)

次に，単位円を x_1 軸方向へ a_1 倍，x_2 軸方向へ a_2 倍に拡大した楕円の方程式を考えてみよう．この方程式を満たすベクトルを $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ とすれば，

$$\frac{y_1^2}{a_1^2} + \frac{y_2^2}{a_2^2} = 1 \tag{16}$$

を満たす．二次形式を使って書けば，

$$\begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} \frac{1}{{a_1}^2} & 0 \\ 0 & \frac{1}{{a_2}^2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 1 \tag{17}$$

となる．ここで行列 Λ を

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} a_1^2 & 0 \\ 0 & a_2^2 \end{pmatrix} \tag{18}$$

とおけば，

$$\mathbf{y}^T \Lambda^{-1} \mathbf{y} = 1 \tag{19}$$

と書くことができる．さて半径 a 円の場合と同じように，この楕円の方程式を写像を使って導出してみよう．$\mathbf{y}$ は単位円を x_1 軸方向へ a_1 倍，x_2 軸方向へ a_2 倍に拡大したものであるから，

$$\mathbf{y} = A\mathbf{x} \tag{20}$$

$$A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix} = \Lambda^{1/2} \tag{21}$$

と書くことができる．なお $A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$ である．この逆写像は

$$A^{-1} = \begin{pmatrix} \frac{1}{a_1} & 0 \\ 0 & \frac{1}{a_2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{pmatrix} = \Lambda^{-1/2} \tag{22}$$

$$\mathbf{x} = A^{-1}\mathbf{y} = \Lambda^{-1/2}\mathbf{x} \tag{23}$$

と書くことができる．さて $\mathbf{x}$ は単位円上であるため $\mathbf{x}^T\mathbf{x} = 1$ を満たす．したがって

$$\mathbf{x}^T\mathbf{x} = 1 \tag{24}$$

$$(A^{-1}\mathbf{y})^T A^{-1}\mathbf{y} = 1 \tag{25}$$

$$\mathbf{y}^T {A^{-1}}^T A^{-1}\mathbf{y} = 1 \tag{26}$$

$$\mathbf{y}^T \Lambda^{-1}\mathbf{y} = 1 \tag{27}$$

となり (19) が得られた.

1.5　単位円から楕円への写像 (2)

次に一般の楕円について考えてみよう．楕円の長軸を a_1，短軸を a_2 とすれば，これは上で述べた楕円を任意の角度だけ回転させたものである．そこで直交変換 P を使って一般の楕円の方程式を導出してみよう（直交変換は図形の向きを変えるだけで，形は合同のまま変わらないことを思い出そう）．したがって任意の楕円は $\mathbf{z} = PA\mathbf{x} = B\mathbf{x}$ として表される．ここで $B = PA$ である．また行列 S を次式で定義する．

$$S \triangleq BB^T \tag{28}$$

$$= (PA)(PA)^T \tag{29}$$

$$= P\Lambda P^T \tag{30}$$

S は実対象行列である．また $S^{-1} = P\Lambda^{-1}P^T$ である．さて $\mathbf{z} = B\mathbf{x}$ の逆変換は

$$\mathbf{x} = B^{-1}\mathbf{z} = (PA)^{-1}\mathbf{z} \tag{31}$$

$$= A^{-1}P^{-1}\mathbf{z} \tag{32}$$

$$= \Lambda^{-1/2}P^T\mathbf{z} \tag{33}$$

と表される．これを (1) に代入すると

$$\mathbf{x}^T\mathbf{x} = 1 \tag{34}$$

$$(B^{-1}\mathbf{z})^T(B^{-1}\mathbf{z}) = 1 \tag{35}$$

$$(\Lambda^{-1/2}P^T\mathbf{z})^T(\Lambda^{-1/2}P^T\mathbf{z}) = 1 \tag{36}$$

$$\mathbf{z}^T P\Lambda^{-1}P^T\mathbf{z} = 1 \tag{37}$$

$$\mathbf{z}^T S^{-1}\mathbf{z} = 1 \tag{38}$$

となる．したがって任意の楕円の方程式は実対称行列 S を使って (38) の式で表される．

■注意　実対称行列 S によって $\mathbf{z}^T S \mathbf{z} = 1$ と表される $\mathbf{z}$ の集合は，楕円だけとは限らない．S によっては双曲面を表現することもある．方程式 $\mathbf{z}^T S \mathbf{z} = 1$ が楕円を表現するには，S の固有値がすべて正でなければならない．すなわち S **が正定値の実対称行列**のとき，方程式 (38) は楕円を表す．固有値に負が含まれる例として，

$$\begin{pmatrix} z_1 & z_2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 1 \tag{39}$$

を考えてみれば良い．なお $S = BB^T$ のように書ける場合は S が正定値行列になることが知られている．

1.6　楕円から単位円への写像

では $\mathbf{z}^T S \mathbf{z} = 1$ を満たす楕円の方程式が与えられたとき，これを単位円に戻すことを考えてみよう．これは $\mathbf{z}$ から $\mathbf{x}$ への逆変換を求めることに相当する．今までの話を復習すると，

$$\mathbf{y} = \Lambda^{\frac{1}{2}} \mathbf{x} \tag{40}$$

$$\mathbf{z} = P\mathbf{y} = P\Lambda^{\frac{1}{2}} \mathbf{x} \tag{41}$$

であった．また

$$\left(\Lambda^{\frac{1}{2}}\right)^{-1} = \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} \end{pmatrix} = \Lambda^{-\frac{1}{2}} \tag{42}$$

$$P^{-1} = P^T \tag{43}$$

であり．逆変換は

$$\mathbf{x} = (P\Lambda^{\frac{1}{2}})^{-1} \mathbf{z} \tag{44}$$

$$= \Lambda^{-\frac{1}{2}} P^T \mathbf{z} \tag{45}$$

であった．

今までは直交行列 P と対角行列 $A^2 = \Lambda$ が与えられたときの楕円の方程式を求めることが目的であった．今度は逆に，楕円の方程式 $\mathbf{z}^T S^{-1} \mathbf{z} = 1$ が与えられているとき，すなわち実対称行列 S が与えられたときに，行列 P, Λ を求める方法を考えてみよう．$S = P\Lambda P^T$ であり，また Λ は対角行列であることから，行列 Λ は行列 S を対角化したものに相当する．したがって λ_1, λ_2 は S の固有値であり，P は固有ベクトルを並べたものになる．以上より次の結論が得られる．

■結論　二次形式 $\mathbf{z}^T S^{-1} \mathbf{z} = 1$ が与えられたとき，線形写像 $\mathbf{x} = \Lambda^{-\frac{1}{2}} P^T \mathbf{z}$ により得られる $\mathbf{x}$ は $\mathbf{x}^T \mathbf{x} = ||\mathbf{x}||^2 = 1$ を満たす．ただし S は正定値実対称行列であり，P は S を対角化する正規直交行列，Λ は S の固有値を並べた対角行列である．

課題

1. 以上の説明を図で表現してみよ．
2. MATLAB を用いて，単位円を任意の楕円に変換してみよ．またその結果を図として出力せよ．
3. MATLAB を用いて正定値実対称行列 S の固有値・固有ベクトルを求めよ．また求めた固有値・固有ベクトルを使って，楕円上の点を単位円上に写像してみよ．

■**注**　MATLAB は高機能関数電卓のようなものであり，行列演算を始め，微分方程式や信号処理などの計算が簡単に行える．MATLAB の基本（たとえば行列計算など）を知るだけならばすぐに覚えられるので，ぜひこの機会に学んでおくと良い．

2　確率論の基礎

本課題で必要となる最低限の確率論について述べる．なお「パターン認識」(R. O. Duda, P. E. Hart and D. G. Stork) の付録 (p.612〜632) に確率論の基礎知識がコンパクトにまとまっているので参照すると良い．

2.1　確率密度関数

サイコロの目のように試行するたびに値が変わるものを**確率変数**と呼ぶ．サイコロの場合は 1,. . . ,6 の値しか取らないため，**離散確率変数**と呼ぶ．確率変数を X とすれば，「$X = 1$ となる確率は 1/6」のように表現できる．これを $\mathrm{Prob}(X = 1) = 1/6$ のように書く．

一方，確率変数が実数値などを取る場合は**連続確率変数**と呼ぶ．連続確率変数の場合，「$X = 3.141592653\ldots$ となる確率」を定義しようにも，限りなくゼロに近くなってしまう．そこで代わりに「密度」を考える．これを**確率密度**と呼ぶ．確率密度は**確率密度関数** $p(x)$ で表されるが，これは x 付近に確率がどれくらい集中しているかを表す．確率密度関数は次式で定義される．

$$p(x) \triangleq \lim_{\Delta x \to 0} \frac{\mathrm{Prob}(x \le X < x + \Delta x)}{\Delta x} \tag{46}$$

（左辺の分子 $\mathrm{Prob}(x \le X < x + \Delta x)$ は，「確率変数 X が，実数 x と $x + \Delta x$ の間に来る確率」と読む）．たとえば九工大生の一人をランダムに選んだときの身長を確率変数 X(cm) とすれば，$p(x)$ は身長 1cm あたりの確率の密度を表す．したがって $p(160)$ は身長が 160〜161cm の範囲に来る確率にほぼ等しい．確率密度関数 $p(x)$ を使うと，確率変数 X がある範囲の値を取る確率を次のようにして求めることができる．

$$\mathrm{Prob}(a \le X < b) = \int_a^b p(x)dx \tag{47}$$

なお「X が何でも良いのでなにかの値を取る確率」は 1，すなわち絶対に起こり得るので，

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \tag{48}$$

である．これは確率密度関数の満たすべき大事な性質である．

離散確率の場合，確率変数を別の値に変換しても確率自体は変わらない．たとえばサイコロの目を $X = 1, 2, 3, \ldots, 6$ の代わりに $Y = 2, 4, 6, \ldots, 12$ と 2 倍の値を割り振ったとしても，（目の数字が変わっただけで）それぞれの目が出る確率に違いはない．すなわち $\mathrm{Prob}(X = 1) = \mathrm{Prob}(Y = 2)$ が成り立つ．しかし連続確率の場合，確率変数を別の値に変換すると，確率密度関数も変化する．たとえば「九工大生をランダムに一人選んだときの身長」を考えた場合，確率変数 X を cm 単位で測った身長，確率変数 Y を mm 単位で測った身長とすると，両者の確率密度 $p_X(x)$ と $p_Y(y) = p_Y(10x)$ は等しくならない．なぜなら $p_X(x)$ は身長 1cm あたりの確率密度，$p_Y(y)$ は身長 1mm あたりの確率密度になるからである．すなわち $p_X(160\mathrm{cm})$ は身長が 160〜161cm の範囲に入る確率とほぼ等しく，$p_Y(1600\mathrm{mm})$ は 1600〜1601mm の範囲に入る確率とほぼ等しい．したがって $p_X(160\mathrm{cm})$ の方が $p_Y(1600\mathrm{mm})$ より約 10 倍大きくなる．

今，$y = f(x)$ という変換（ただし $f(x)$ は単調増加もしくは単調減少関数）を行ったとき，y の確率密度関数 $q(y)$ は $p(x)$ を使って次のように表される．

$$q(y) = \left|\frac{\partial x}{\partial y}\right| p(f^{-1}(y)) \tag{49}$$

特に線形変換 $y = ax$ に対しては,

$$q(y) = \left|\frac{1}{a}\right| p\left(\frac{y}{a}\right) \tag{50}$$

となる．

2.2　期待値と平均

連続確率変数 X の期待値は確率密度関数を用いて次式のように表される．

$$E[x] = \int_{-\infty}^{+\infty} xp(x)dx \tag{51}$$

$E[x]$ を特に**平均**と呼び，しばしば $E[x] = \mu$ と書く．また確率変数の関数 $f(x)$ の期待値は次式のように表される．

$$E[f(x)] = \int_{-\infty}^{+\infty} f(x)\, p(x)\, dx \tag{52}$$

期待値を求める計算は線形性があるため,

$$E[a_1 f_1(x) + a_2 f_2(x)] = a_1 E[f_1(x)] + a_2 E[f_2(x)] \tag{53}$$

が成り立つ．特に定数の期待値は,

$$E[C] = C \tag{54}$$

となる．

実際に確率変数を何度も観測して得られた数値の集合を**サンプル集合と呼ぶ**．たとえばサイコロを実際に 100 回振れば，100 個の数値を得ることができる．今，サンプル集合を $X = \{x_1, \ldots, x_N\}$ とすれば，サンプル集合の**アンサンブル平均**（もしくは単にサンプル平均）

$$\hat{\mu} = \frac{x_1 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{55}$$

が得られる．アンサンブル平均もまた確率変数であるが，サンプル数を十分大きく取れば平均値 μ に近づく．これを**大数の法則**と言う．

$$\langle x \rangle = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i = \mu \tag{56}$$

したがってサンプル集合のアンサンブル平均 $\hat{\mu}$ は平均値 μ の推定値として使うことができる．

2.3　分散と標準偏差

確率変数 X の分布のばらつきを表すの量が**分散**と**標準偏差**である．分散は期待値を使って次式で定義される．

$$Var[x] = E[(x-\mu)^2] = \int_{-\infty}^{+\infty} (x-\mu)^2 p(x) dx \tag{57}$$

分散の平方根 $\sqrt{Var[x]}$ が標準偏差であり，しばしば σ で表す．分散はばらつき幅の二乗に比例するので，ばらつき幅そのものを表すには標準偏差の方が便利である．

サンプル集合 $X = \{x_1, \ldots, x_N\}$ から分散を推定するには，次式を用いる．

$$\hat{\sigma}^2 = \frac{(x_1-\hat{\mu})^2 + \cdots + (x_N-\hat{\mu})^2}{N-1} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\hat{\mu})^2 \tag{58}$$

平均値推定の場合と異なり，分母が $(N-1)$ であることに注意する．

2.4　多次元（ベクトル）確率変数

確率変数がベクトルであっても上記の議論は同様に成り立つ．たとえばランダムに九工大生を一人選んだときに得られる (身長, 体重) ベクトルはその例である．SOM で扱う多次元データも本質的にはベクトル確率変数とみなされる．ここでは話をわかりやすくするため 2 次元ベクトル $\mathbf{x} = (x_1, x_2)$ の場合について考えるが，3 次元以上であっても同様である．

まず確率密度関数 $p(\mathbf{x}) = p(x_1, x_2)$ は次式で与えられる．

$$p(\mathbf{x}) = p(x_1, x_2) = \lim_{\Delta x_1, \Delta x_2 \to 0} \frac{\mathrm{Prob}\left((x_1 \le X_1 < x_1 + \Delta x_1) \ \wedge \ (x_2 \le X_2 < x_2 + \Delta x_2)\right)}{\Delta x_1 \Delta x_2} \tag{59}$$

$\mathbf{y} = f(\mathbf{x})$ という変換を行った場合（ただし $f(\mathbf{x})$ は x_1,x_2 双方について単調関数とする），$\mathbf{y}$ の確率密度 $q(\mathbf{y})$ はヤコビアン J を使って次のように表される．

$$q(\mathbf{y}) = \frac{1}{|J|} p(f^{-1}(\mathbf{y})) \tag{60}$$

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix} \tag{61}$$

特に線形変換 $\mathbf{y} = A\mathbf{x}$, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ の場合は，$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ より

$$\frac{\partial y_1}{\partial x_1} = \frac{\partial(ax_1 + bx_2)}{\partial x_1} = a \tag{62}$$

$$\frac{\partial y_1}{\partial x_2} = \frac{\partial(ax_1 + bx_2)}{\partial x_2} = b \tag{63}$$

$$\frac{\partial y_2}{\partial x_1} = \frac{\partial(cx_1 + dx_2)}{\partial x_1} = c \tag{64}$$

$$\frac{\partial y_2}{\partial x_2} = \frac{\partial(cx_1 + dx_2)}{\partial x_2} = d \tag{65}$$

$$\tag{66}$$

より,

$$J = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det A \tag{67}$$

となるため,

$$q(\mathbf{y}) = \frac{p(A^{-1}\mathbf{y})}{|\det A|} \tag{68}$$

となる.

■注意　$\det A$ は行列 A の行列式 $|A|$ を意味するが，絶対値記号との混乱を避けるため，ここでは $\det A$ の方を用いた.

2.5　多次元確率変数の平均と分散

確率変数がベクトルであっても，期待値を同様に定義できる.

$$\mu = \int_{\mathbf{x}\in\mathbb{R}^2} \mathbf{x}p(\mathbf{x})d\mathbf{x} \tag{69}$$

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \iint_{-\infty}^{+\infty} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} p(x_1, x_2)dx_1dx_2 \tag{70}$$

また確率変数の関数についても同様である.

$$E[f(\mathbf{x})] = \int_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{71}$$

サンプル集合 $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ が与えられた場合，平均の推定値 $\hat{\mu}$ はアンサンブル平均

$$\hat{\mu} = \frac{\mathbf{x}_1 + \ldots \mathbf{x}_N}{N} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i \tag{72}$$

で与えられる.

分散は $s_{11} = E[(x_1 - \mu_1)^2]$, $s_{22} = E[(x_2 - \mu_2)^2]$ の他に $s_{12} = s_{21} = E[(x_1 - \mu_1)(x_2 - \mu_2)]$ がある．これを x_1, x_2 の**共分散**と呼ぶ．また次の行列 S を**共分散行列**と呼ぶ.

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} = \begin{pmatrix} E[(x_1 - \mu_1)^2] & E[(x_1 - \mu_1)(x_2 - \mu_2)] \\ E[(x_1 - \mu_1)(x_2 - \mu_2)] & E[(x_2 - \mu_2)^2] \end{pmatrix} \tag{73}$$

$$= E\left[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T\right] \tag{74}$$

特に $\mu = 0$ の場合，$S = E[\mathbf{x}\mathbf{x}^T]$ となる．定義より共分散行列は実対称行列であり，また正定値行列であることが知られている．すなわち共分散行列は直交行列により対角化でき，固有値はすべて正（もしくは非負）である.

サンプル集合から共分散行列を推定する場合は,

$$\hat{S} = \frac{(\mathbf{x}_1 - \mu)(\mathbf{x}_1 - \mu)^T + \ldots (\mathbf{x}_N - \mu)(\mathbf{x}_N - \mu)^T}{N-1} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T \tag{75}$$

となる.

2.6 無相関と独立

x_1, x_2 の共分散が $s_{12} > 0$ のとき，x_1 が増加/減少すると x_2 も増加/減少する傾向がある．このとき，**x_1 と x_2 は正の相関がある**と言う．逆に $s_{12} < 0$ のときは x_1 が増加/減少すると x_2 は逆に減少/増加する傾向がある．このとき x_1, x_2 は負の相関があると言う．$s_{12} = 0$ のときは**無相関**であると言う．s_{12} が無相関の場合，共分散行列 S は対角行列になる．s_{12} の大きさは x_1, x_2 間の相関およびそれぞれの分散の双方で決まる．そこで $\rho = s_{12}/(s_{11}s_{12})$ とすれば，ρ は必ず $-1 \leq \rho \leq +1$ の範囲の値を取る．この ρ を**相関係数**と呼ぶ.

一方，確率密度関数が $p(\mathbf{x}) = p(x_1, x_2) = p_1(x_1)p_2(x_2)$ のように積で書ける場合，x_1 と x_2 は**独立**であると言う．

無相関と独立は似た概念であるが，無相関は単に増減の関連性のみで決まるのに対し，独立は確率密度全体の形状で決まるため，独立の方が強い（制約が厳しい）概念である．したがって独立であれば必ず無相関であると言える（確率密度が x_1, x_2 個別に決まれば，増減関係も無関係であるといえる）．しかし無相関だからといって独立であるとは限らない（x_1, x_2 の増減関係に傾向が見られなくても，確率密度関数としては独立にならないことがある）．ただし確率変数が後に述べるガウス分布に従うとわかっている場合は，無相関すなわち独立であると言える．

与えられたデータ集合を無相関化する解析法を主成分分析と言い，与えられたデータを独立化する解析法を独立成分分析と言う．

3 ガウス分布と主成分分析

3.1 1 次元標準正規分布

もっとも重要な確率密度関数として，**ガウス分布**がある（正規分布とも呼ぶ）．パターン認識の多くの問題では，データがガウス分布に従うと仮定して扱うことが多い（本当にガウス分布に従うかどうかはチェックすべきであるが）．

平均 $\mu = 0$, 分散 $\sigma^2 = 1$ のガウス分布を**標準正規分布**と呼ぶ．標準正規分布の確率密度は

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x^2}{2}\right] \tag{76}$$

で表される．右辺の $\frac{1}{\sqrt{2\pi}}$ は $p(x)$ の積分が 1 になるための係数である．したがって

$$p(x) \propto \exp\left[-\frac{x^2}{2}\right] \tag{77}$$

と考えればわかりやすい（$\exp(-x^2/2)$ を区間 $(-\infty, +\infty)$ で積分すると $\sqrt{2\pi}$ になる）．両辺の対数をとると

$$-\log p(x) = \log \frac{1}{p(x)} = \frac{1}{2}x^2 + const \tag{78}$$

となる．

3.2　1 次元ガウス分布

次に平均 μ，分散 σ^2 の 1 変数のガウス分布の確率密度関数を導出してみよう．そのためには $y = \sigma x + \mu$ と変数変換すればよい．逆変換は

$$x = \frac{y-\mu}{\sigma} \tag{79}$$

であり，また

$$\frac{dy}{dx} = \sigma \tag{80}$$

であるから，

$$q(y) = \frac{1}{\sigma} p\left(\frac{y-\mu}{\sigma}\right) \tag{81}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right] \tag{82}$$

となる．これが一般の場合の 1 次元ガウス分布の確率密度である．

3.3　2 次元ガウス分布

今，確率変数 $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ を考える．x_1 と x_2 はそれぞれ独立で，かつそれぞれ 1 次元標準正規分布に従うとする．すなわち

$$p(\mathbf{x}) = p(x_1, x_2) = p_1(x_1)p_2(x_2) \tag{83}$$

$$p(x_1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x_1^2}{2}\right] \tag{84}$$

$$p(x_2) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{x_2^2}{2}\right] \tag{85}$$

$$\tag{86}$$

であるとする．このとき $\mathbf{x}$ の確率密度関数 $p(\mathbf{x})$ は

$$p(\mathbf{x}) = \frac{1}{2\pi} \exp\left[-\frac{(x_1^2 + y_2^2)}{2}\right] \tag{87}$$

$$= \frac{1}{2\pi} \exp\left[-\frac{\mathbf{x}^T\mathbf{x}}{2}\right] \tag{88}$$

となる．これが 2 次元の標準正規分布の確率密度関数である．

さて，ここで $\mathbf{y} = A\mathbf{x} + \mu$ という変数変換を考える．逆変換は

$$\mathbf{x} = A^{-1}(\mathbf{y} - \mu) \tag{89}$$

であり，またヤコビアン $J = |A|$ であるから，

$$p(\mathbf{y}) = \frac{1}{2\pi|A|} \exp\left[-\frac{1}{2}\left(A^{-1}(\mathbf{y}-\mu)\right)^T\left(A^{-1}(\mathbf{y}-\mu)\right)\right] \tag{90}$$

となる．さてここで $S = AA^T$ とおくと，S は次の式を満たす．

$$|S| = |A|^2 \tag{91}$$

$$S^{-1} = (A^T)^{-1}A^{-1} \tag{92}$$

したがって

$$p(\mathbf{y}) = \frac{1}{2\pi|S|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{y}-\mu)S^{-1}(\mathbf{y}-\mu)\right] \tag{93}$$

となる．

■**課題**　S は $\mathbf{y}$ の共分散行列であることを示せ．

3.4　D 次元ガウス分布

D 次元の場合も話は同様である．D 次元の標準正規分布は

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}}\exp\left[-\frac{\|\mathbf{x}\|^2}{2}\right] \tag{94}$$

で与えられる．これは 1 次元の標準正規分布を D 回乗じたものになっている．ここで $\mathbf{y} = A\mathbf{x} + \mu$ という線形変換を行うと，次式が成り立つ．

$$\mathbf{x} = A^{-1}(\mathbf{y}-\mu) \tag{95}$$

$$E[\mathbf{y}] = \mu \tag{96}$$

$$E[(\mathbf{y}-\mu)(\mathbf{y}-\mu)^T] = E[(A\mathbf{x})(A\mathbf{x})^T] = AA^T = S \tag{97}$$

$$J = |A| = |S|^{1/2} \tag{98}$$

したがって平均が μ, 共分散が $S = AA^T$ となる．このときの $\mathbf{y}$ の確率密度関数は

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|S|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{y}-\mu)^T S^{-1}(\mathbf{y}-\mu)\right] \tag{99}$$

で与えられる．これが D 次元のガウス分布の一般形である．

3.5　主成分分析

今，$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ は 2 次元標準正規分布に従う 2 次元確率変数とする．すなわち確率密度関数が (88) で与えられるとする（$\mathbf{x}$ の共分散行列 S_x は単位行列 I であること，また平均 $\mu_x = 0$ であることに注意）．また確率変数 $\mathbf{y}$ は対角行列 $A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$ によって $\mathbf{y} = A\mathbf{x}$ と変換されたものとする．すなわち

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{100}$$

である．このとき，$\mathbf{y}$ の共分散行列 S_y を求めてみよう．S_y を求めるのに $\mathbf{y}$ の確率密度関数は必要ない．(71) に $\mathbf{y} = A\mathbf{x}$ を入れれば良い．したがって

$$
\begin{aligned}
S_y &= E\left[\mathbf{y}\mathbf{y}^T\right] &(101)\\
&= E\left[A\mathbf{x}(A\mathbf{x})^T\right] &(102)\\
&= E\left[A\mathbf{x}\mathbf{x}^T A^T\right] &(103)\\
&= E\left[A(\mathbf{x}\mathbf{x}^T)A^T\right] &(104)\\
&= AE\left[\mathbf{x}\mathbf{x}^T\right]A^T &(105)\\
&= AS_xA^T &(106)\\
&= AIA^T = AA^T = A^2 = \begin{pmatrix} a_1^2 & 0 \\ 0 & a_2^2 \end{pmatrix} &(107)
\end{aligned}
$$

ここで $S_y = A^2 = \Lambda$ とおけば，$A = \Lambda^{1/2}$ であり，また $\mathbf{x} = \Lambda^{-1/2}\mathbf{y}$ である．

次に $\mathbf{y}$ を直交行列 P によって回転し，原点を μ だけずらしたものを $\mathbf{z} = P\mathbf{y} + \mu$ としよう．すなわち $\mathbf{z} = PA\mathbf{x} = P\Lambda^{1/2}\mathbf{x} + \mu$ である．また P は直交行列なので $P^{-1} = P^T$ であることも思い出そう．このとき，$\mathbf{z}$ の共分散行列 S_z は次のようになる．

$$
\begin{aligned}
S_z &= E\left[(\mathbf{z}-\mu)(\mathbf{z}-\mu)^T\right] &(108)\\
&= E\left[(PA\mathbf{x})(PA\mathbf{x})^T\right] &(109)\\
&= E\left[PA\mathbf{x}\mathbf{x}^T A^T P^T\right] &(110)\\
&= PAE\left[\mathbf{x}\mathbf{x}^T\right]A^T P^T &(111)\\
&= PAA^T P^T = P\Lambda P^T &(112)
\end{aligned}
$$

したがって S_z を対角化したものが Λ であり，P は対角化行列である．すなわち λ_1, λ_2 は $\mathbf{z}$ の共分散行列 S_z の固有値である．

さて，現実の問題では S_z をサンプル集合 $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ から求めることができるが，それに対応する $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ はわからない．しかし $\mathbf{x} = \Lambda^{-1/2}P^T\mathbf{z}$ であるから，Λ と P がわかれば $\mathbf{x}_i$ を求めることができる．すなわち次の手順を踏むことで $\mathbf{z}$ のサンプル集合 $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$ から隠された真の姿 $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ を知ることができる．

1. サンプル集合 Z から $\mathbf{z}$ の平均 μ および共分散行列 S_z を求める．
2. S_z の固有値・固有ベクトルを求める．
3. 固有値から Λ, 固有ベクトルから P を求める．なお固有値は通常，値の大きい順に並べる．
4. $\mathbf{y}_i = P^T(\mathbf{z}_i - \mu), \mathbf{x}_i = \Lambda^{-1/2}P^T(\mathbf{z}_i - \mu)$ と変換する．

これを主成分分析と言う．またこの手続きによって共分散行列が対角化されるため，「白色化」「無相関化」と呼ぶこともある．主成分分析では，目的に応じて $\mathbf{y}$ を求める場合と，$\mathbf{x}$ を求める場合の 2 とおりある．

課題

1. D 次元標準正規分布に従う $\mathbf{x}$ のサンプル集合を生成せよ．これを $X = \{\mathbf{x}_i\}$ とする．
2. 適当な変換行列 A, P とベクトル μ を用意して，$\mathbf{z}_i = PA\mathbf{x}_i + \mu$ と変換し，$\mathbf{z}$ のサンプル集合を生成せよ．

3. $\mathbf{z}$ のサンプル集合から平均の推定値 $\hat{\mu}$ および共分散行列の推定値 $\hat{S}_z$ を求めてみよ． またそれは理論値（データ生成時に実際に用いた μ および $S_z = (PA)(PA)^T$）とどれくらい違うか比較してみよ．
4. $\hat{S}_z$ の固有値・固有ベクトルを求め，主成分分析を行ってみよ．
5. $\hat{S}_z$ から求めた固有値・固有ベクトル $\hat{\Lambda}, \hat{P}$ を用いて，$\{\mathbf{x}_i\}$ の推定をせよ．すなわち $\hat{\mathbf{x}}_i = \hat{\Lambda}^{-1/2} P^T(\mathbf{z}_i - \hat{\mu})$ と変換を行い，元の $\{\mathbf{x}_i\}$ とどれくらい違うか比較してみよ．
6. 2 つの固有値が同じとき，すなわち $\lambda_1 = \lambda_2$ のときは，$\mathbf{x}_i$ を復元することができない．なぜか．