

## 1 分散共分散行列

分散共分散行列

$d$  次元特徴量ベクトル  $\{\mathbf{x}_i\}$   $i = 1, \dots, n$  のデータに対して、分散共分散行列  $\Sigma$  は次式で定義される

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (1)$$

$$= E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}]^T \quad (2)$$

ある特徴量縦ベクトルを考える。

$$\mathbf{x} = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^d \end{pmatrix} \quad (3)$$

これらの特徴量のデータセット  $\{\mathbf{x}_i\}$   $i = 1, \dots, n$  があつた場合に、各次元に対して分散は次式で計算できる：

$$Cov[x^i, x^i] = E[(x^i - E[x^i])^2] = E[(x^i)^2] - E[x^i]^2 \quad (4)$$

具体的に書き出すと

$$V[x^i] = \frac{1}{n} \sum_{k=1}^n (x_k^{(i)} - \bar{x}^i)^2 \quad (5)$$

となる。また次元  $x_i$  と  $x_j$  の共分散は次式で定義される：

$$Cov[x^i, x^j] = E[(x^i - E[x^i])(x^j - E[x^j])] = E[x^i x^j] - E[x^i]E[x^j] \quad (6)$$

具体的に書き出すと

$$Cov[x^i, x^j] = \frac{1}{n} \sum_{k=1}^n (x_k^{(i)} - \bar{x}^i)(x_k^{(j)} - \bar{x}^j) \quad (7)$$

となる。ここで添字  $k$  は特徴量ベクトルの個数 (= データ数) を表しており、データ数と次元数の添字を混同しないこと。

以上を分散共分散行列  $S$  としてまとめたい。まず

$$\mathbf{x}\mathbf{x}^T = \begin{pmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{pmatrix} (x^1 \ x^2 \ \dots \ x^d) = \begin{pmatrix} (x^1)^2 & x^1 x^2 & \dots & x^1 x^n \\ x^2 x^1 & \dots & \dots & x^2 x^n \\ \vdots & \vdots & \vdots & \vdots \\ x^d x^1 & \dots & \dots & (x^d)^2 \end{pmatrix} \quad (8)$$

であり、

$$E[\mathbf{x}\mathbf{x}^T] = \begin{pmatrix} E[(x^1)^2] & E[x^1 x^2] & \dots & E[x^1 x^d] \\ E[x^2 x^1] & \dots & \dots & E[x^2 x^d] \\ \vdots & \vdots & \vdots & \vdots \\ E[x^d x^1] & \dots & \dots & E[(x^d)^2] \end{pmatrix} \quad (9)$$

$$E[\mathbf{x}]E[\mathbf{x}]^T = \begin{pmatrix} E[x^1]^2 & E[x^1]E[x^2] & \dots & E[x^1]E[x^d] \\ E[x^2]E[x^1] & \dots & \dots & E[x^2]E[x^d] \\ \vdots & \vdots & \vdots & \vdots \\ E[x^d]E[x^1] & \dots & \dots & E[x^d]^2 \end{pmatrix} \quad (10)$$

である。以上の  $E[\mathbf{x}\mathbf{x}^T]$  と  $E[\mathbf{x}]E[\mathbf{x}]^T$  を用いると分散共分散行列は

$$\Sigma = \begin{pmatrix} Cov[x^1, x^1] & \dots & Cov[x^1, x^d] \\ Cov[x^2, x^1] & \dots & Cov[x^2, x^d] \\ \vdots & \vdots & \vdots \\ Cov[x^n, x^1] & \dots & Cov[x^n, x^d] \end{pmatrix} = E[\mathbf{x}\mathbf{x}^T] - E[\mathbf{x}]E[\mathbf{x}]^T \quad (11)$$

と表され、具体的に書き出すと

$$\Sigma = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T - \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \right) \left( \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^T \right) \quad (12)$$

と表される<sup>\*1</sup>。

## 2 Decorrelation

<https://courses.media.mit.edu/2010fall/mas622j/whiten.pdf>

共分散行列は

$$\Sigma = E(\mathbf{x}\mathbf{x}^T) - E(\mathbf{x})E(\mathbf{x})^T \quad (13)$$

で表される。特徴量データを  $\mathbf{x}' = \mathbf{x} - E(\mathbf{x})$  と平均ゼロとなるように変換すると、

$$\Sigma = E(\mathbf{x}'(\mathbf{x}')^T) \quad (14)$$

で表すことができる。簡単のために変換後の特徴量ベクトルを  $\mathbf{x}$  と表す。decorrelation とは相関係数をゼロにする（無相関にする）操作である。相関係数は共分散を用いることで

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (15)$$

と計算されることを思い出すと、相関係数がゼロとは、共分散  $\sigma_{xy}$  がゼロになることと同値である。つまり分散共分散行列の対角成分だけが残りでそれ以外がゼロとなるような変換をすれば良いことになる。これは分散共分散行列を対角化することに等しく、固有値分解をすることで解くことができる。

---

\*1 期待値の計算  $E[\cdot]$  で和の計算を表していることに留意