PCA procedure

1. The given data can be written as an $m \times n$ matrix.

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix}$$

2. Construct an $n \times n$ covarance matrix as follows:

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & d_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

   where

$$c_{ij} = cov(d_i, d_j) = \frac{1}{n-1} \sum_{k=1}^{m} (d_{ki} - \mu_{d_i})(d_{kj} - \mu_{d_j})$$

   and

$$\mu_{d_i} = \frac{1}{n} \sum_{k=1}^{m} d_{ki}, \quad \mu_{dj} = \frac{1}{n} \sum_{k=1}^{m} d_{kj},$$

   The covarance matrix is a symmetric matrix.

3. Compute the eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ and the corresponding eigenvectors $\{e_1, e_2, ..., e_n\}$. Keep the larger eigenvalues and its eigenvector and ignore the smaller eigenvalues.

4. Choose the larger eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_p\}$ and eigenvectors $\{e_1, e_2, ..., e_p\}$ to reduce the dimension from n to p. We obtain the following principal component of eigenvectors

$$E = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1p} \\ e_{21} & e_{22} & \cdots & e_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ e_{n1} & e_{n2} & \cdots & e_{np} \end{bmatrix}$$

5. The matrix after dimension reduction can be expressed as follows:

$$F = (E^T \widehat{D}^T)^T$$

where

$$\widehat{D} = \begin{bmatrix} d_{11} - \mu_{d_1} & d_{12} - \mu_{d_2} & \cdots & d_{1n} - \mu_{d_n} \\ d_{21} - \mu_{d_1} & d_{22} - \mu_{d_2} & \cdots & d_{2n} - \mu_{d_n} \\ \cdots & \cdots & \cdots & \cdots \\ d_{m1} - \mu_{d_1} & d_{n2} - \mu_{d_2} & \cdots & d_{mn} - \mu_{d_n} \end{bmatrix}$$

F is the data after dimension reduction. The dimension of F is $m \times p$.

Example 1 $A = \begin{pmatrix} 2.6 & 1.4 & 2.8 \\ 0.5 & 0.7 & 0.3 \\ 2.2 & 2.9 & 1.4 \\ 1.9 & 2.2 & 0.9 \\ 3.1 & 3.0 & 2.1 \\ 2.3 & 2.7 & 1.6 \\ 2.0 & 1.6 & 0.4 \\ 1.0 & 1.1 & 0.5 \\ 1.5 & 1.6 & 2.3 \\ 1.3 & 0.9 & 1.7 \end{pmatrix}$

B=repmat(mean(A),10,1);
$\widehat{D} = A - B$;
C=cov($\widehat{D}$);
[eigenvectors, eigenvalues]=eig(C);
E=[eigenvectors(:,2), eigenvectors(:,3)];
$F = (E^T \widehat{D}^T)^T$

$$F = \begin{pmatrix} -1.2730 & 0.9681 \\ 0.0585 & -2.0552 \\ 0.6946 & 0.8587 \\ 0.6368 & 0.0008 \\ 0.2976 & 1.8414 \\ 0.4268 & 0.9091 \\ 0.6822 & -0.5521 \\ 0.1947 & -1.4077 \\ -0.8752 & 0.1429 \\ -0.8431 & -0.7061 \end{pmatrix}$$