Data Science
Unit-II

Principal Component Analysis (PCA)

Dimensionality Reduction

From : Web resources

Dimensionality Reduction:

In machine learning classification problems, there are often too many factors on the basis of which the final classification is done. These factors are basically variables called features. The higher the number of features, the harder it gets to visualize the training set and then work on it. Sometimes, most of these features are correlated, and hence redundant. This is where dimensionality reduction algorithms come into play.

Dimensionality reduction is the process of reducing the number of random variables under consideration, by obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

•Problem with high dimensional data?

- It can mean high computational cost to perform learning.
- It often leads to over-fitting when learning a model, which means that the model will perform well on the training data but poorly on test data.
- Data are rarely randomly distributed in high-dimensions and are highly correlated, often with spurious correlations.
- The distances between a nearest and farthest data point can become equidistant in high dimensions, that can hamper the accuracy of some distance-based analysis tools.

Importance of Dimensionality reduction?

•It reduces the time and storage space required.
•It helps Remove multi-collinearity which improves the interpretation of the parameters of the machine learning model.
•It becomes easier to visualize the data when reduced to very low dimensions such as 2D or 3D.
•It avoids the curse of dimensionality.
•It removes irrelevant features from the data, Because having irrelevant features in the data can decrease the accuracy of the models and make your model learn based on irrelevant features.

Methods for Dimensionality Reduction:

1. Principal Component Analysis (PCA)
2. Multi Dimensionality Scaling (MDS)

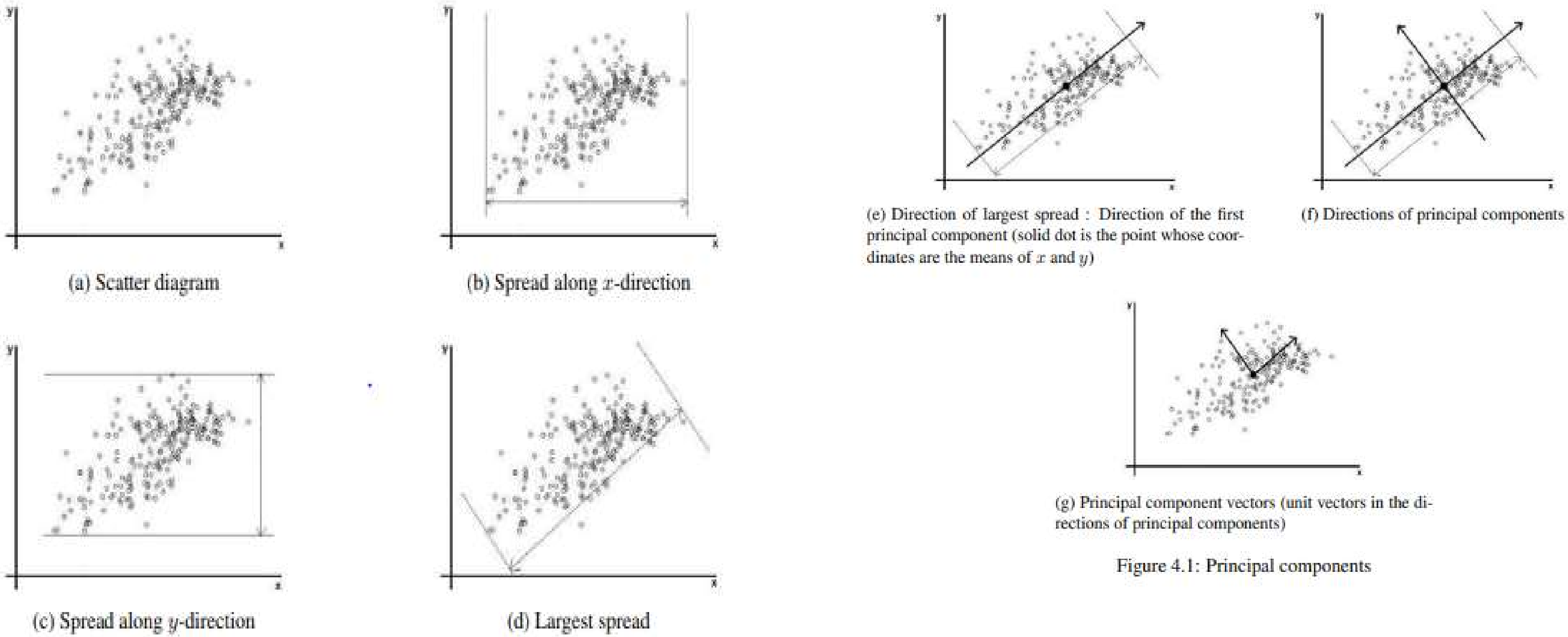
Introduction to PCA

- Principal Components Analysis, or PCA, is a data analysis tool that is usually used to reduce the dimensionality (number of variables) of a large number of interrelated variables, while retaining as much of the information (variation) as possible
- PCA calculates an uncorrelated set of variables (components or pc's). These components are ordered so that the first few retain most of the variation present in all of the original variables.
- Unlike its cousin Factor Analysis, PCA always yields the same solution from the same data (apart from arbitrary differences in the sign).
- The computations of PCA reduce to an eigenvalue-eigenvector problem. PCA is eigen vector of maximum eigen value
- Note that **PCA is a data analytical, rather than statistical, procedure. Hence, you will not find many t-tests or F-tests in PCA.**

Introduction to PCA

- Principal component analysis (PCA) is a **statistical procedure** that uses an **orthogonal transformation** to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.
- **The number of principal components is less than or equal to the smaller of the number of original variables or the number of observations.**
- This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components.

Graphical illustration of the idea



(a) Scatter diagram

(b) Spread along x-direction

(c) Spread along y-direction

(d) Largest spread

(e) Direction of largest spread : Direction of the first principal component (solid dot is the point whose coordinates are the means of x and y)

(f) Directions of principal components

(g) Principal component vectors (unit vectors in the directions of principal components)

Figure 4.1: Principal components

Graphical illustration of the idea

Let us examine the figures in Figure 4.1.

(i) Figure 4.1a shows a scatter diagram of a two-dimensional data.

(ii) Figure 4.1b shows spread of the data in the x direction and Figure 4.1c shows the spread of the data in the y-direction. We note that the spread in the x-direction is more than the spread in the y direction.

(iii) Examining Figures 4.1d and 4.1e, we note that the maximum spread occurs in the direction shown in Figure 4.1e. Figure 4.1e also shows the point whose coordinates are the mean values of the two features in the dataset. This direction is called the *direction of the first principal component* of the given dataset.

(iv) The direction which is perpendicular (orthogonal) to the direction of the first principal component is called the *direction of the second principal component* of the dataset. This direction is shown in Figure 4.1f. (This is only with reference to a two-dimensional dataset.)

(v) The unit vectors along the directions of principal components are called the *principal component vectors*, or simply, *principal components*. These are shown in Figure 4.1g.

PCA Algorithm

Step 1. Data

We consider a dataset having n features or variables denoted by $X_1, X_2, \ldots, X_n$. Let there be N examples. Let the values of the i-th feature X_i be $X_{i1}, X_{i2}, \ldots, X_{iN}$ (see Table 4.1).

Features	Example 1	Example 2	$\cdots$	Example N
X_1	X_{11}	X_{12}	$\cdots$	X_{1N}
X_2	X_{21}	X_{22}	$\cdots$	X_{2N}
$\vdots$				
X_i	X_{i1}	X_{i2}	$\cdots$	X_{iN}
$\vdots$				
X_n	X_{n1}	X_{n2}	$\cdots$	X_{nN}

Table 4.1: Data for PCA algorithm

Step 2. Compute the means of the variables

We compute the mean $\bar{X}_i$ of the variable X_i:

$$\bar{X}_i = \frac{1}{N}(X_{i1} + X_{i2} + \cdots + X_{iN}).$$

PCA Algorithm

Step 3. Calculate the covariance matrix

Consider the variables X_i and X_j (i and j need not be different). The covariance of the ordered pair (X_i, X_j) is defined as[1]

$$\operatorname{Cov}(X_i, X_j) = \frac{1}{N-1}\sum_{k=1}^{N}(X_{ik} - \bar{X}_i)(X_{jk} - \bar{X}_j). \tag{4.1}$$

We calculate the following $n \times n$ matrix S called the covariance matrix of the data. The element in the i-th row j-th column is the covariance $\operatorname{Cov}(X_i, X_j)$:

$$S = \begin{bmatrix} \operatorname{Cov}(X_1, X_1) & \operatorname{Cov}(X_1, X_2) & \cdots & \operatorname{Cov}(X_1, X_n) \\ \operatorname{Cov}(X_2, X_1) & \operatorname{Cov}(X_2, X_2) & \cdots & \operatorname{Cov}(X_2, X_n) \\ \vdots & & & \\ \operatorname{Cov}(X_n, X_1) & \operatorname{Cov}(X_n, X_2) & \cdots & \operatorname{Cov}(X_n, X_n) \end{bmatrix}$$

PCA Algorithm

Step 4. Calculate the eigenvalues and eigenvectors of the covariance matrix

Let S be the covariance matrix and let I be the identity matrix having the same dimension as the dimension of S.

i) Set up the equation:

$$\det(S - \lambda I) = 0. \tag{4.2}$$

This is a polynomial equation of degree n in λ. It has n real roots (some of the roots may be repeated) and these roots are the eigenvalues of S. We find the n roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of Eq. (4.2).

ii) If $\lambda = \lambda'$ is an eigenvalue, then the corresponding eigenvector is a vector

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

such that

$$(S - \lambda' I)U = 0.$$

(This is a system of n homogeneous linear equations in u_1, u_2, ..., u_n and it always has a nontrivial solution.) We next find a set of n orthogonal eigenvectors $U_1, U_2, \ldots, U_n$ such that U_i is an eigenvector corresponding to λ_i.[2]

PCA Algorithm

iii) We now normalise the eigenvectors. Given any vector X we normalise it by dividing X by its length. The length (or, the norm) of the vector

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is defined as

$$||X|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Given any eigenvector U, the corresponding normalised eigenvector is computed as

$$\frac{1}{||U||} U.$$

We compute the n normalised eigenvectors $e_1, e_2, \ldots, e_n$ by

$$e_i = \frac{1}{||U_i||} U_i, \quad i = 1, 2, \ldots, n.$$

PCA Algorithm

Step 5. Derive new data set

Order the eigenvalues from highest to lowest. The unit eigenvector corresponding to the largest eigenvalue is the first principal component. The unit eigenvector corresponding to the next highest eigenvalue is the second principal component, and so on.

i) Let the eigenvalues in descending order be $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ and let the corresponding unit eigenvectors be $e_1, e_2, \ldots, e_n$.

ii) Choose a positive integer p such that $1 \leq p \leq n$.

iii) Choose the eigenvectors corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ and form the following $p \times n$ matrix (we write the eigenvectors as row vectors):

$$F = \begin{bmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_p^T \end{bmatrix},$$

where T in the superscript denotes the transpose.

PCA Algorithm

iv) We form the following $n \times N$ matrix:

$$X = \begin{bmatrix} X_{11} - \bar{X}_1 & X_{12} - \bar{X}_1 & \cdots & X_{1N} - \bar{X}_1 \\ X_{21} - \bar{X}_2 & X_{22} - \bar{X}_2 & \cdots & X_{2N} - \bar{X}_2 \\ \vdots & & & \\ X_{n1} - \bar{X}_n & X_{n2} - \bar{X}_n & \cdots & X_{nN} - \bar{X}_n \end{bmatrix}$$

v) Next compute the matrix:

$$X_{\text{new}} = FX.$$

Note that this is a $p \times N$ matrix. This gives us a dataset of N samples having p features.

Step 6. New dataset

The matrix X_{new} is the new dataset. Each row of this matrix represents the values of a feature. Since there are only p rows, the new dataset has only features.

Step 7. Conclusion

This is how the principal component analysis helps us in dimensional reduction of the dataset. Note that it is not possible to get back the original n-dimensional dataset from the new dataset.

PCA: Example

Problem

Given the data in Table 4.2, use PCA to reduce the dimension from 2 to 1.

Feature	Example 1	Example 2	Example 3	Example 4
X_1	4	8	13	7
X_2	11	4	5	14

Table 4.2: Data for illustrating PCA

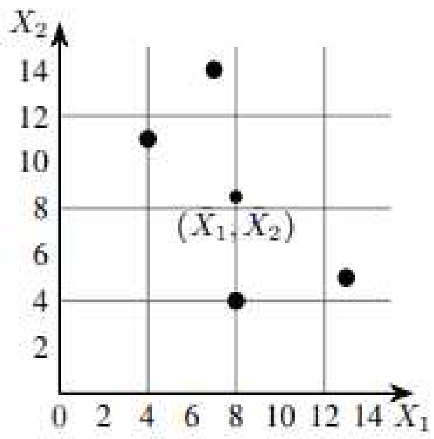


Figure 4.2: Scatter plot of data in Table 4.2

Solution

1. Scatter plot of data

We have

$$\bar{X}_1 = \frac{1}{4}(4 + 8 + 13 + 7) = 8,$$
$$\bar{X}_2 = \frac{1}{4}(11 + 4 + 5 + 14) = 8.5.$$

Figure 4.2 shows the scatter plot of the data together with the point $(\bar{X}_1, \bar{X}_2)$.

Example

2. Calculation of the covariance matrix

The covariances are calculated as follows:

$$
\begin{aligned}
\mathrm{Cov}(X_1, X_2) &= \frac{1}{N-1}\sum_{k=1}^{N}(X_{1k} - \bar{X}_1)^2 \\
&= \tfrac{1}{3}\left((4-8)^2 + (8-8)^2 + (13-8)^2 + (7-8)^2\right) \\
&= 14 \\
\mathrm{Cov}(X_1, X_2) &= \frac{1}{N-1}\sum_{k=1}^{N}(X_{1k} - \bar{X}_1)(X_{2k} - \bar{X}_2) \\
&= \tfrac{1}{3}((4-8)(11-8.5) + (8-8)(4-8.5) \\
&\quad + (13-8)(5-8.5) + (7-8)(14-8.5) \\
&= -11 \\
\mathrm{Cov}(X_2, X_1) &= \mathrm{Cov}(X_1, X_2) \\
&= -11 \\
\mathrm{Cov}(X_2, X_2) &= \frac{1}{N-1}\sum_{k=1}^{N}(X_{2k} - \bar{X}_2)^2 \\
&= \tfrac{1}{3}\left((11-8.5)^2 + (4-8.5)^2 + (5-8.5)^2 + (14-8.5)^2\right) \\
&= 23
\end{aligned}
$$

The covariance matrix is

$$
\begin{aligned}
S &= \begin{bmatrix} \mathrm{Cov}(X_1, X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Cov}(X_2, X_2) \end{bmatrix} \\
&= \begin{bmatrix} 14 & -11 \\ -11 & 23 \end{bmatrix}
\end{aligned}
$$

Example

3. Eigenvalues of the covariance matrix

The characteristic equation of the covariance matrix is

$$
\begin{aligned}
0 &= \det(S - \lambda I) \\
&= \begin{vmatrix} 14-\lambda & -11 \\ -11 & 23-\lambda \end{vmatrix} \\
&= (14-\lambda)(23-\lambda) - (-11)\times(-11) \\
&= \lambda^2 - 37\lambda + 201
\end{aligned}
$$

Solving the characteristic equation we get

$$
\begin{aligned}
\lambda &= \tfrac{1}{2}(37 \pm \sqrt{565}) \\
&= 30.3849,\ 6.6151 \\
&= \lambda_1,\ \lambda_2 \quad \text{(say)}
\end{aligned}
$$

Example

4. Computation of the eigenvectors

To find the first principal components, we need only compute the eigenvector corresponding to the largest eigenvalue. In the present example, the largest eigenvalue is λ_1 and so we compute the eigenvector corresponding to λ_1.

The eigenvector corresponding to $\lambda = \lambda_1$ is a vector $U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ satisfying the following equation:

$$\begin{aligned}
\begin{bmatrix} 0 \\ 0 \end{bmatrix} &= (S - \lambda_1 I)X \\
&= \begin{bmatrix} 14 - \lambda_1 & -11 \\ -11 & 23 - \lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\
&= \begin{bmatrix} (14 - \lambda_1)u_1 - 11u_2 \\ -11u_1 + (23 - \lambda_1)u_2 \end{bmatrix}
\end{aligned}$$

This is equivalent to the following two equations:

$$\begin{aligned}
(14 - \lambda_1)u_1 - 11u_2 &= 0 \\
-11u_1 + (23 - \lambda_1)u_2 &= 0
\end{aligned}$$

Using the theory of systems of linear equations, we note that these equations are not independent and solutions are given by

$$\frac{u_1}{11} = \frac{u_2}{14 - \lambda_1} = t,$$

that is

$$u_1 = 11t, \quad u_2 = (14 - \lambda_1)t,$$

Example

where t is any real number. Taking $t = 1$, we get an eigenvector corresponding to λ_1 as

$$U_1 = \begin{bmatrix} 11 \\ 14 - \lambda_1 \end{bmatrix}.$$

To find a unit eigenvector, we compute the length of X_1 which is given by

$$\begin{aligned} ||U_1|| &= \sqrt{11^2 + (14 - \lambda_1)^2} \\ &= \sqrt{11^2 + (14 - 30.3849)^2} \\ &= 19.7348 \end{aligned}$$

Therefore, a unit eigenvector corresponding to $lambda_1$ is

$$\begin{aligned} e_1 &= \begin{bmatrix} 11/||U_1|| \\ (14 - \lambda_1)/||U_1|| \end{bmatrix} \\ &= \begin{bmatrix} 11/19.7348 \\ (14 - 30.3849)/19.7348 \end{bmatrix} \\ &= \begin{bmatrix} 0.5574 \\ -0.8303 \end{bmatrix} \end{aligned}$$

By carrying out similar computations, the unit eigenvector e_2 corresponding to the eigenvalue $\lambda = \lambda_2$ can be shown to be

$$e_2 = \begin{bmatrix} 0.8303 \\ 0.5574 \end{bmatrix}.$$

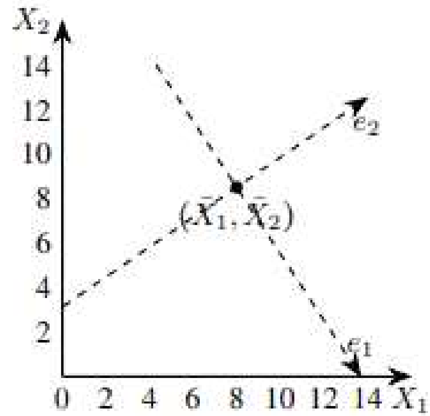


Figure 4.3: Coordinate system for principal components

Example

5. Computation of first principal components

Let $\begin{bmatrix} X_{1k} \\ X_{2k} \end{bmatrix}$ be the k-th sample in Table 4.2. The first principal component of this example is given by (here "T" denotes the transpose of the matrix)

$$e_1^T \begin{bmatrix} X_{1k} - \bar{X}_1 \\ X_{2k} - \bar{X}_2 \end{bmatrix} = \begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} X_{1k} - \bar{X}_1 \\ X_{2k} - \bar{X}_2 \end{bmatrix}$$
$$= 0.5574(X_{1k} - \bar{X}_1) - 0.8303(X_{2k} - \bar{X}_2).$$

For example, the first principal component corresponding to the first example $\begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \end{bmatrix}$ is calculated as follows:

$$\begin{bmatrix} 0.5574 & -0.8303 \end{bmatrix} \begin{bmatrix} X_{11} - \bar{X}_1 \\ X_{21} - \bar{X}_2 \end{bmatrix} = 0.5574(X_{11} - \bar{X}_1) - 0.8303(X_{21} - \bar{X}_2)$$
$$= 0.5574(4 - 8) - 0.8303(11 - 8,5)$$
$$= -4.30535$$

The results of calculations are summarised in Table 4.3.

X_1	4	8	13	7
X_2	11	4	5	14
First principal components	-4.3052	3.7361	5.6928	-5.1238

Table 4.3: First principal components for data in Table 4.2

Example

6. Geometrical meaning of first principal components

As we have seen in Figure 4.1, we introduce new coordinate axes. First we shift the origin to the "center" $(\bar{X}_1, \bar{X}_2)$ and then change the directions of coordinate axes to the directions of the eigenvectors e_1 and e_2 (see Figure 4.3).

Next, we drop perpendiculars from the given data points to the e_1-axis (see Figure 4.4). The first principal components are the e_1-coordinates of the feet of perpendiculars, that is, the projections on the e_1-axis. The projections of the data points on e_1-axis may be taken as approximations of the given data points hence we may replace the given data set with these points. Now, each of these

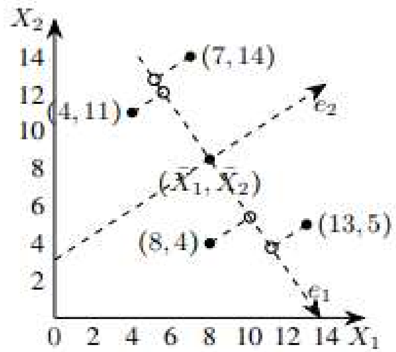


Figure 4.4: Projections of data points on the axis of the first principal component

PC1 components	-4.305187	3.736129	5.692828	-5.123769

Table 4.4: One-dimensional approximation to the data in Table 4.2

Example

approximations can be unambiguously specified by a single number, namely, the e_1-coordinate of approximation. Thus the two-dimensional data set given in Table 4.2 can be represented approximately by the following one-dimensional data set (see Figure 4.5):

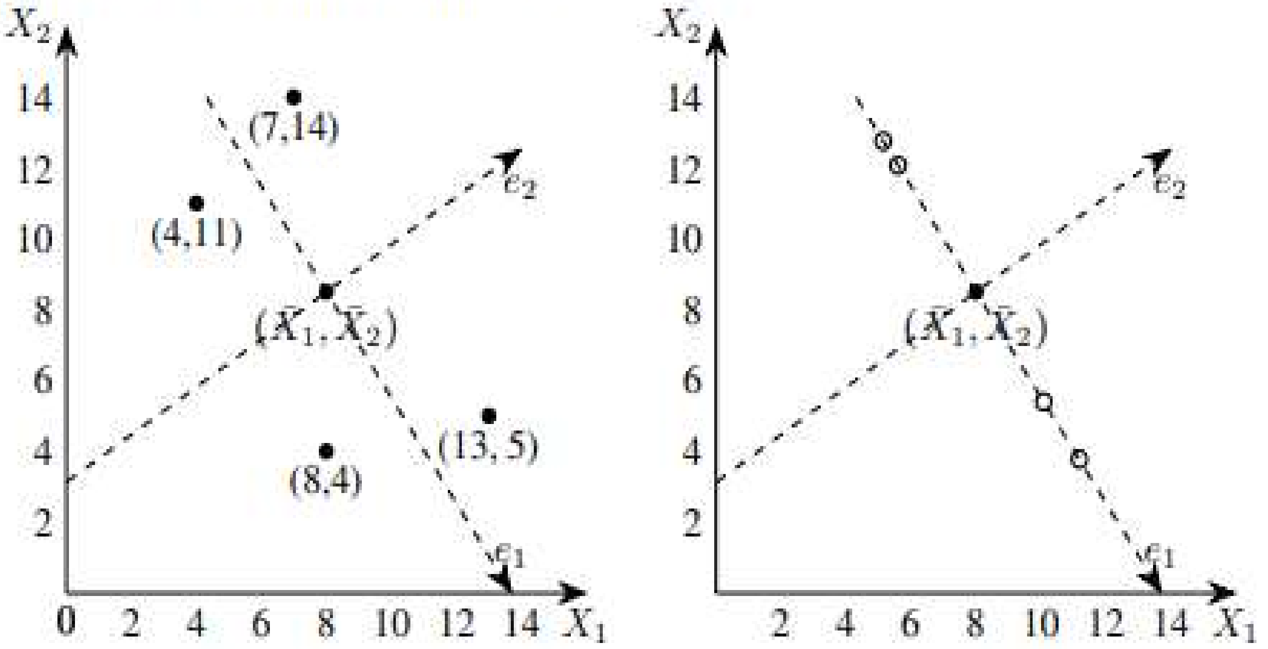


Figure 4.5: Geometrical representation of one-dimensional approximation to the data in Table 4.2

PCA : Example 2 : Find PCA s for given students data

Step 1: Take the whole dataset consisting of *d+1 dimensions* and ignore the labels such that our new dataset becomes *d dimensional.*

Let our data matrix **X** be the score of three students :

Student	Math	English	Art
1	90	60	90
2	90	90	30
3	60	60	60
4	60	60	90
5	30	30	30

Example: Step 2 **Compute the mean of every dimension of the whole dataset.**

$$\mathbf{A} = \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix}$$

Step 2. Compute the means of the variables

We compute the mean $\bar{X}_i$ of the variable X_i:

$$\bar{X}_i = \frac{1}{N}(X_{i1} + X_{i2} + \cdots + X_{iN}).$$

So, The mean of matrix **A** would be

$$\bar{\mathbf{A}} = [\ 66\ \ 60\ \ 60\ ]$$

Step **3. Compute the *covariance matrix* of the whole dataset (sometimes also called as the variance-covariance matrix)**

we can compute the covariance of two variables **X** and **Y** using the following formula

$$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n} (Xi - \bar{x})(Yi - \bar{y})$$

$$\begin{array}{c|c} & \begin{array}{ccc} Math & English & Arts \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{bmatrix} 90 & 60 & 90 \\ 90 & 90 & 30 \\ 60 & 60 & 60 \\ 60 & 60 & 90 \\ 30 & 30 & 30 \end{bmatrix} \end{array}$$

Its *covariance matrix* would be

$$\begin{array}{c|c} & \begin{array}{ccc} Math & English & Art \end{array} \\ \begin{array}{c} Math \\ English \\ Art \end{array} & \begin{bmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{bmatrix} \end{array}$$

Step 4. Compute Eigenvectors and corresponding Eigenvalues

Let **A** be a square matrix, **v** a vector and **λ** a scalar that satisfies **A**v = **λ**v, then **λ** is called eigenvalue associated with eigenvector **v** of **A**.
The eigenvalues of **A** are roots of the characteristic equation

$$\det(A - \lambda I) = 0$$

Calculating *det(A-λI)* first, *I* is an identity matrix :

$$\det\left(\begin{pmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)$$

$$\begin{pmatrix} 504 & 360 & 180 \\ 360 & 360 & 0 \\ 180 & 0 & 720 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix}$$

$$\det\begin{pmatrix} 504-\lambda & 360 & 180 \\ 360 & 360-\lambda & 0 \\ 180 & 0 & 720-\lambda \end{pmatrix}$$

$$-\lambda^3 + 1584\lambda^2 - 641520\lambda + 25660800$$

$$-\lambda^3 + 1584\lambda^2 - 641520\lambda + 25660800 = 0$$

$$\lambda \approx 44.81966\ldots, \lambda \approx 629.11039\ldots, \lambda \approx 910.06995\ldots$$

$$\begin{pmatrix} -3.75100\ldots \\ 4.28441\ldots \\ 1 \end{pmatrix}, \begin{pmatrix} -0.50494\ldots \\ -0.67548\ldots \\ 1 \end{pmatrix}, \begin{pmatrix} 1.05594\ldots \\ 0.69108\ldots \\ 1 \end{pmatrix}$$

Step 5. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a *d × k dimensional* matrix W.

after sorting the eigenvalues in decreasing order, we have

$$\begin{pmatrix} 910.06995 \\ 629.11039 \\ 44.81966 \end{pmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 1.05594 & -0.50494 \\ 0.69108 & -0.67548 \\ 1 & 1 \end{bmatrix}$$

Step 6. Transform the samples onto the new subspace

In the last step, we use the 2×3 dimensional matrix ***W*** that we just computed to transform our samples onto the new subspace via the equation ***y = W' × x*** where ***W'*** is the *transpose* of the matrix ***W.***

PCA: Example: Find PCAs for given data set

Consider the two dimensional patterns (2, 1), (3, 5), (4, 3), (5, 6), (6, 7), (7, 8).

Compute the principal component using PCA Algorithm.

OR

Compute the principal component of following data-

CLASS 1

$X = 2 , 3 , 4$

$Y = 1 , 5 , 3$

CLASS 2

$X = 5 , 6 , 7$

$Y = 6 , 7 , 8$

Step-01:

Get data.

The given feature vectors are-

- $x_1 = (2, 1)$
- $x_2 = (3, 5)$
- $x_3 = (4, 3)$
- $x_4 = (5, 6)$
- $x_5 = (6, 7)$
- $x_6 = (7, 8)$

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \end{bmatrix} \begin{bmatrix} 6 \\ 7 \end{bmatrix} \begin{bmatrix} 7 \\ 8 \end{bmatrix}$$

Example

Step-02:

Calculate the mean vector (μ).

Mean vector (μ)

= ((2 + 3 + 4 + 5 + 6 + 7) / 6, (1 + 5 + 3 + 6 + 7 + 8) / 6)

= (4.5, 5)

Thus,

$$\textbf{Mean vector } (\mu) = \begin{bmatrix} 4.5 \\ 5 \end{bmatrix}$$

Step-03:

Subtract mean vector (μ) from the given feature vectors.

- $x_1 - \mu = (2 - 4.5, 1 - 5) = (-2.5, -4)$
- $x_2 - \mu = (3 - 4.5, 5 - 5) = (-1.5, 0)$
- $x_3 - \mu = (4 - 4.5, 3 - 5) = (-0.5, -2)$
- $x_4 - \mu = (5 - 4.5, 6 - 5) = (0.5, 1)$
- $x_5 - \mu = (6 - 4.5, 7 - 5) = (1.5, 2)$
- $x_6 - \mu = (7 - 4.5, 8 - 5) = (2.5, 3)$

Feature vectors (x_i) after subtracting mean vector (μ) are-

$$\begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} \begin{bmatrix} -0.5 \\ -2 \end{bmatrix} \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} 2.5 \\ 3 \end{bmatrix}$$

Example: Covariance matrix method

Step-04:

Calculate the covariance matrix.

Covariance matrix is given by-

$$\text{Covariance Matrix} = \frac{\sum (x_i - \mu)(x_i - \mu)^t}{n}$$

Now,

$$m_1 = (x_1 - \mu)(x_1 - \mu)^t = \begin{bmatrix} -2.5 \\ -4 \end{bmatrix} \begin{bmatrix} -2.5 & -4 \end{bmatrix} = \begin{bmatrix} 6.25 & 10 \\ 10 & 16 \end{bmatrix}$$

$$m_2 = (x_2 - \mu)(x_2 - \mu)^t = \begin{bmatrix} -1.5 \\ 0 \end{bmatrix} \begin{bmatrix} -1.5 & 0 \end{bmatrix} = \begin{bmatrix} 2.25 & 0 \\ 0 & 0 \end{bmatrix}$$

$$m_3 = (x_3 - \mu)(x_3 - \mu)^t = \begin{bmatrix} -0.5 \\ -2 \end{bmatrix} \begin{bmatrix} -0.5 & -2 \end{bmatrix} = \begin{bmatrix} 0.25 & 1 \\ 1 & 4 \end{bmatrix}$$

$$m_4 = (x_4 - \mu)(x_4 - \mu)^t = \begin{bmatrix} 0.5 \\ 1 \end{bmatrix} \begin{bmatrix} 0.5 & 1 \end{bmatrix} = \begin{bmatrix} 0.25 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

$$m_5 = (x_5 - \mu)(x_5 - \mu)^t = \begin{bmatrix} 1.5 \\ 2 \end{bmatrix} \begin{bmatrix} 1.5 & 2 \end{bmatrix} = \begin{bmatrix} 2.25 & 3 \\ 3 & 4 \end{bmatrix}$$

$$m_6 = (x_6 - \mu)(x_6 - \mu)^t = \begin{bmatrix} 2.5 \\ 3 \end{bmatrix} \begin{bmatrix} 2.5 & 3 \end{bmatrix} = \begin{bmatrix} 6.25 & 7.5 \\ 7.5 & 9 \end{bmatrix}$$

Example

Now,

Covariance matrix

$= (m_1 + m_2 + m_3 + m_4 + m_5 + m_6) / 6$

On adding the above matrices and dividing by 6, we get-

$$\text{Covariance Matrix} = \frac{1}{6}\begin{bmatrix} 17.5 & 22 \\ 22 & 34 \end{bmatrix}$$

$$\text{Covariance Matrix} = \begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix}$$

Step-05:

Calculate the eigen values and eigen vectors of the covariance matrix.

λ is an eigen value for a matrix M if it is a solution of the characteristic equation $|M - \lambda I| = 0$.

So, we have-

$$\left|\begin{matrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{matrix}\right| - \left|\begin{matrix} \lambda & 0 \\ 0 & \lambda \end{matrix}\right| = 0$$

$$\left|\begin{matrix} 2.92 - \lambda & 3.67 \\ 3.67 & 5.67 - \lambda \end{matrix}\right| = 0$$

Example

From here,

$(2.92 - \lambda)(5.67 - \lambda) - (3.67 \times 3.67) = 0$

$16.56 - 2.92\lambda - 5.67\lambda + \lambda^2 - 13.47 = 0$

$\lambda^2 - 8.59\lambda + 3.09 = 0$

Solving this quadratic equation, we get $\lambda = 8.22, 0.38$

Thus, two eigen values are $\lambda_1 = 8.22$ and $\lambda_2 = 0.38$.

Clearly, the second eigen value is very small compared to the first eigen value.

So, the second eigen vector can be left out.

Eigen vector corresponding to the greatest eigen value is the principal component for the given data set.

So. we find the eigen vector corresponding to eigen value λ_1.

We use the following equation to find the eigen vector-

$$MX = \lambda X$$

where-

- M = Covariance Matrix
- X = Eigen vector
- λ = Eigen value

Substituting the values in the above equation, we get-

$$\begin{bmatrix} 2.92 & 3.67 \\ 3.67 & 5.67 \end{bmatrix} \begin{bmatrix} X1 \\ X2 \end{bmatrix} = 8.22 \begin{bmatrix} X1 \\ X2 \end{bmatrix}$$

Example

Solving these, we get-

$2.92X_1 + 3.67X_2 = 8.22X_1$

$3.67X_1 + 5.67X_2 = 8.22X_2$

On simplification, we get-

$5.3X_1 = 3.67X_2$ ………(1)

$3.67X_1 = 2.55X_2$ ………(2)

From (1) and (2), $\mathbf{X_1 = 0.69X_2}$

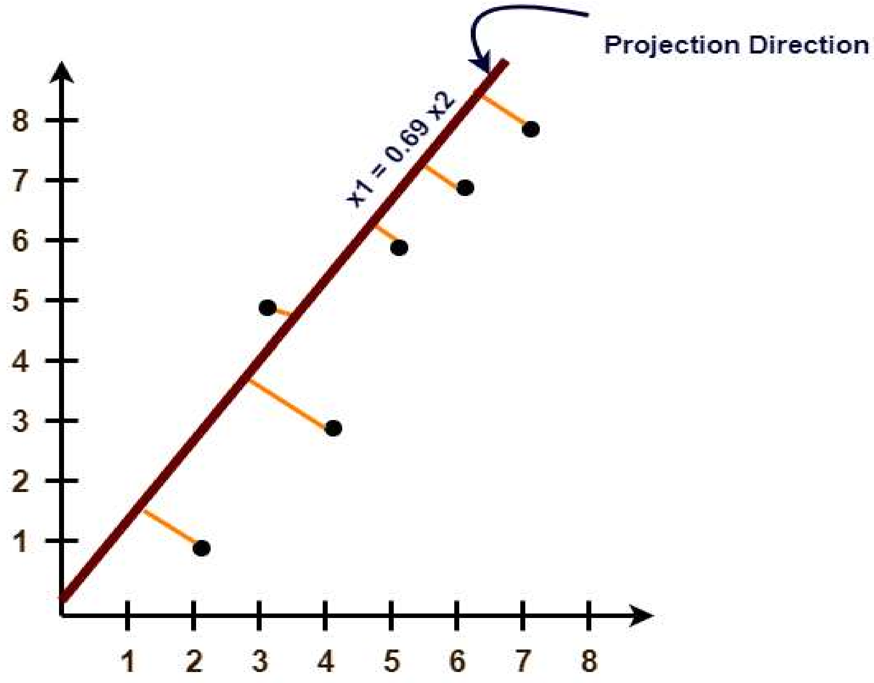
From (2), the eigen vector is-

Eigen Vector : $\begin{bmatrix} X1 \\ X2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$

Thus, principal component for the given data set is-

Principal Component : $\begin{bmatrix} X1 \\ X2 \end{bmatrix} = \begin{bmatrix} 2.55 \\ 3.67 \end{bmatrix}$

Lastly, we project the data points onto the new s



Difference between PCA and MDS

	PCA	MDS
Input data	Data matrix (S subjects in G dimensions)	Dissimilarity structure (distance between any pair of subjects)
Method	"Project" subjects to low-dimensional space and preserve as large variance as possible	Find a low-dimensional space that best keep the dissimilarity structure
Restrictions	Data have to be in Euclidean space	Flexible to any data structure as long as the dissimilarity structure can be defined
Pros and cons	The PCs can be further used to model in downstream analyses. If a new subject is added, it can be similarly projected.	Flexibility and visualization. But if a new subject is added, it can't be shown in an existing MDS solution.

Exercise

1. Describe the forward selection algorithm for implementing the subset selection procedure for dimensionality reduction.
2. Describe the backward selection algorithm for implementing the subset selection procedure for dimensionality reduction.
3. What is the first principal component of a data? How one can compute it?
4. Describe with the use of diagrams the basic principle of PCA.
5. Explain the procedure for the computation of the principal components of a given data.
6. Describe how principal component analysis is carried out to reduce dimensionality of data sets.
7. Given the following data, compute the principal component vectors and the first principal components:

x	2	3	7
y	11	14	26

8. Given the following data, compute the principal component vectors and the first principal components:

x	-3	1	-2
y	2	-1	3