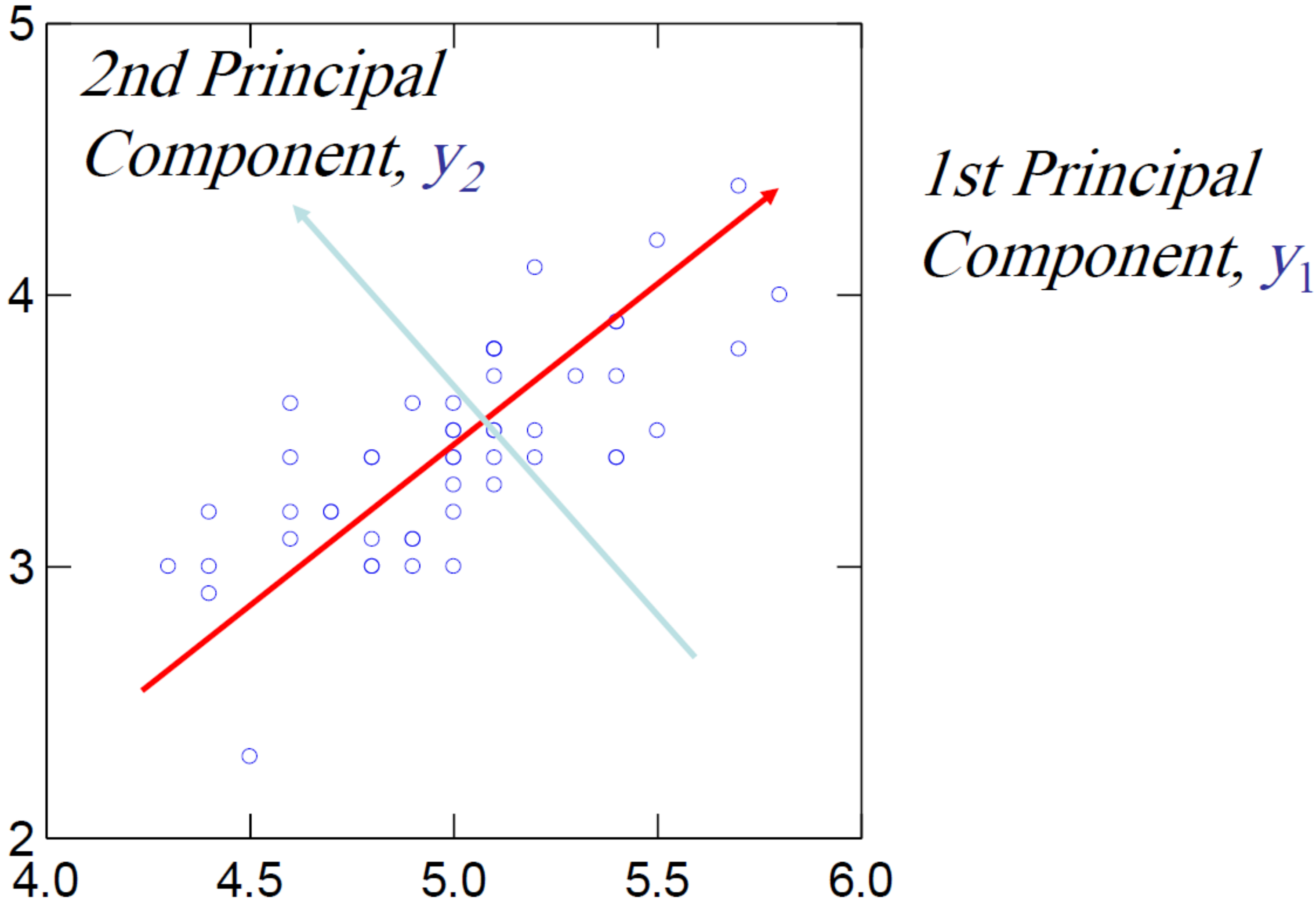
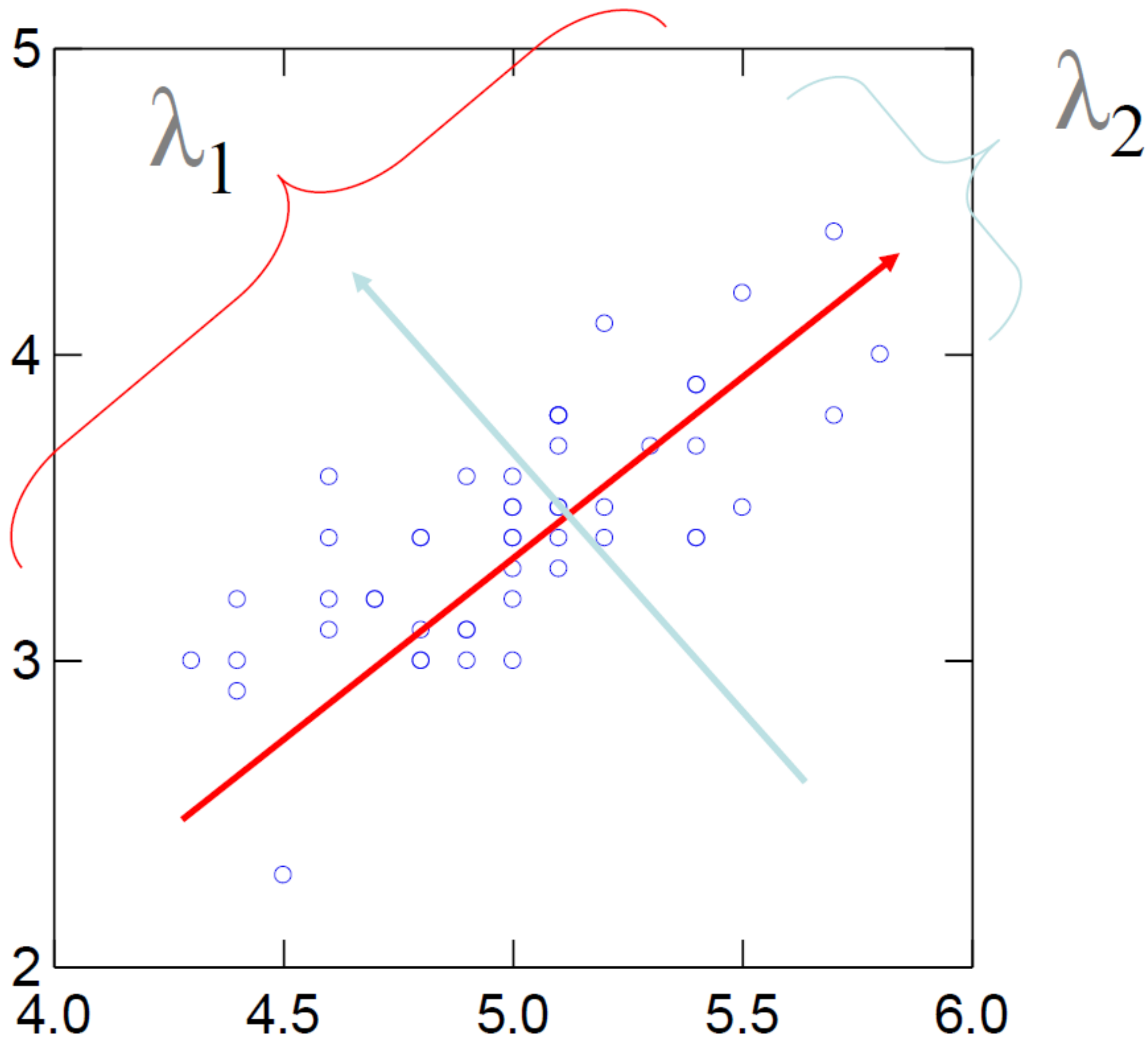

2nd Principal Component, y_2
1st Principal Component, y_1
5
4
3
2
4.0
4.5
5.0
5.5
6.0

PCA Eigenvalues



Dataset

Consider the following dataset

x1	**2.5**	**0.5**	**2.2**	**1.9**	**3.1**	**2.3**	**2.0**	**1.0**	**1.5**	**1.1**
x2	**2.4**	**0.7**	**2.9**	**2.2**	**3.0**	**2.7**	**1.6**	**1.1**	**1.6**	**0.9**

Step 1: Standardize the Dataset

Mean for x_1= 1.81 = x_{1mean}

Mean for x_2= 1.91 = x_{2mean}

We will change the dataset.

We will change the dataset.

$x_{1new} = x_1 - x_{1mean}$	0.69	-1.31	0.39	0.09	1.29	0.49	0.19	-0.81	-0.31	-0.71
$x_{2new} = x_2 - x_{2mean}$	0.49	-1.21	0.99	0.29	1.09	0.79	-0.31	-0.81	-0.31	-1.01

Step 2: Find the Eigenvalues and eigenvectors

Correlation Matrix c = $C = \left(\frac{X \cdot X^T}{N-1}\right)$

$$C = \begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix}$$

Compute Eigenvalues and Eigenvectors

$$\begin{bmatrix} 0.616556 & 0.615444 \\ 0.615444 & 0.716556 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{vmatrix} 0.616556 - \lambda & 0.615444 \\ 0.615444 & 0.716556 - \lambda \end{vmatrix} = 0$$

We get two values for λ, that are **(λ_1) = 1.28403 and (λ_2) = 0.0490834**.

So now we found the eigenvectors for the eigenvector λ_1, they are 0.67787 and 0.73518

So now we found the eigenvectors for the eigenvector \lambda_2, they are 0.735176 and 0.677873

Step 4: Form Feature Vector

$$\begin{bmatrix} 0.677873 & 0.735179 \\ 0.735179 & -0.677879 \end{bmatrix}$$ This is the FEATURE VECTOR for Numerical

Where first column are the eigenvectors of λ_1 & second column are the eigenvectors of λ_2

Step 5: Transform Original Dataset

Use the equation Z = X V

$$Z \begin{bmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{bmatrix} \cdot \begin{bmatrix} 0.677873 & 0.735179 \\ 0.735179 & -0.677879 \end{bmatrix} = \begin{bmatrix} 0.8297008 & 0.17511574 \\ -1.77758022 & -0.14285816 \\ 0.99219768 & -0.38437446 \\ 0.27421048 & -0.13041706 \\ 1.67580128 & 0.20949934 \\ 0.91294918 & -0.17528196 \\ -1.14457212 & -0.04641786 \\ -0.43804612 & -0.01776486 \\ -1.22382.62 & 0.16267464 \end{bmatrix} =$$

Step 6: Reconstructing Data

Use the equation X = $Z * V^T$ ($V^T$is Transpose of V), X = Row Zero Mean Data

$$\begin{bmatrix} 0.8297008 & 0.17511574 \\ -1.77758022 & -0.14285816 \\ 0.99219768 & -0.38437446 \\ 0.27421048 & -0.13041706 \\ 1.67580128 & 0.20949934 \\ 0.91294918 & -0.17528196 \\ -1.14457212 & -0.04641786 \\ -0.43804612 & -0.01776486 \\ -1.22382.62 & 0.16267464 \end{bmatrix} \cdot \begin{bmatrix} 0.677873 & 0.735179 \\ 0.735176 & -0.677879 \end{bmatrix} = \begin{bmatrix} 0.68999997766573 & 0.48999998834233 \\ -1.30999995556827 & -1.20999995590657 \\ 0.389999968063 & 0.98999999665083 \\ 0.08999999969553 & 0.28999999901893 \\ 0.61212695653593 & 0.35482096313253 \\ 0.48999998834233 & 0.78999999732743 \\ 0.189999935723 & -0.309999995127 \\ -0.80999999725977 & -0.80999999725977 \\ -0.30999999895127 & -0.30999999895127 \\ -0.70999999759807 & -1.00999999658317 \end{bmatrix}$$

So in order to reconstruct the original data, we follow:

Row Original DataSet = Row Zero Mean Data + Original Mean

$$\begin{bmatrix} 0.68999997766573 & 0.48999998834233 \\ -1.30999995556827 & -1.20999995590657 \\ 0.389999968063 & 0.98999999665083 \\ 0.08999999969553 & 0.28999999901893 \\ 0.61212695653593 & 0.35482096313253 \\ 0.48999998834233 & 0.78999999732743 \\ 0.189999935723 & -0.309999995127 \\ -0.80999999725977 & -0.80999999725977 \\ -0.30999999895127 & -0.30999999895127 \\ -0.70999999759807 & -1.00999999658317 \end{bmatrix} + \begin{bmatrix} 1.81 & 1.91 \end{bmatrix} = \begin{bmatrix} 2.49 & 2.39 \\ 0.5 & 0.7 \\ 2.19 & 2.89 \\ 1.89 & 2.19 \\ 3.08 & 2.99 \\ 2.30 & 2.7 \\ 2.01 & 1.59 \\ 1.01 & 1.11 \\ 1.5 & 1.6 \\ 1.1 & 0.9 \end{bmatrix}$$

Principal Component Analysis

- In PCA, a new set of features are extracted from the original features which are quite dissimilar in nature. So, an **n-dimensional feature space** gets transformed into an **m-dimensional feature space.**, where the dimensions are orthogonal to each other.

Working of PCA:

- First, calculate the covariance matrix of a data set.
- Then, calculate the eigenvectors of the covariance matrix.
- The eigenvector having the highest eigenvalue represents the direction in which there is the highest variance.
- The eigenvector having the next highest eigenvalue represents the direction in which data has the highest remaining variance and also orthogonal to the first direction. So, this helps in identifying the second principal component.
- Like this, identify the top ‘k’ eigenvectors having top ‘k’ eigenvalues to get the ‘k’ principal components.

Applications of PCA Analysis

- PCA in machine learning is used to visualize multidimensional data.
- In healthcare data to explore the factors that are assumed to be very important in increasing the risk of any chronic disease.
- PCA helps to resize an image.
- PCA is used to analyze stock data and forecasting data.
- You can also use Principal Component Analysis to analyze patterns when you are dealing with high-dimensional data sets.