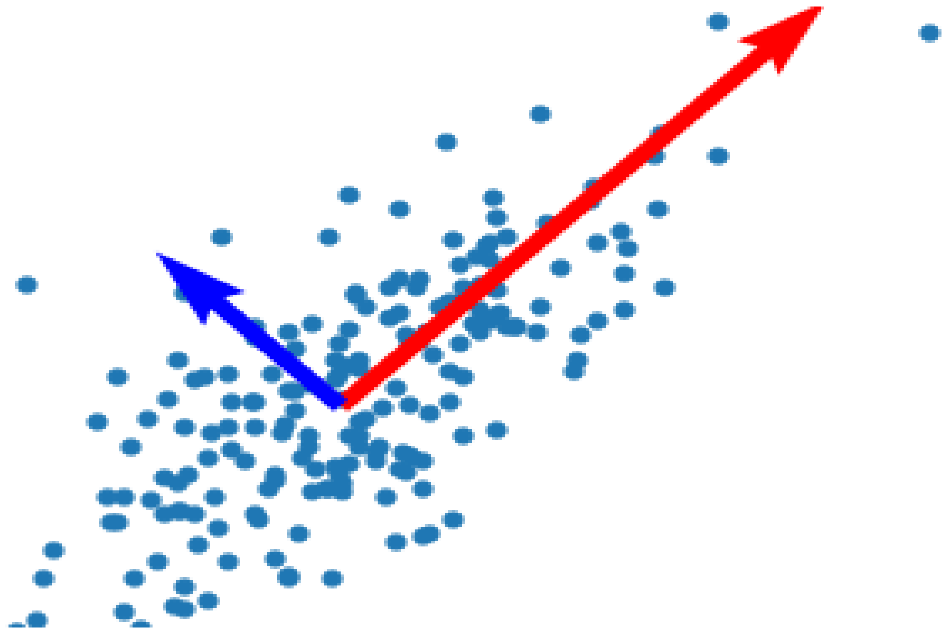
Hochschule Karlsruhe
University of
Applied Sciences

Fakultät für
Elektro- und
Informationstechnik

Linear Principal Components Analysis



Kawther Aboalam

Principal Components Analysis (PCA)

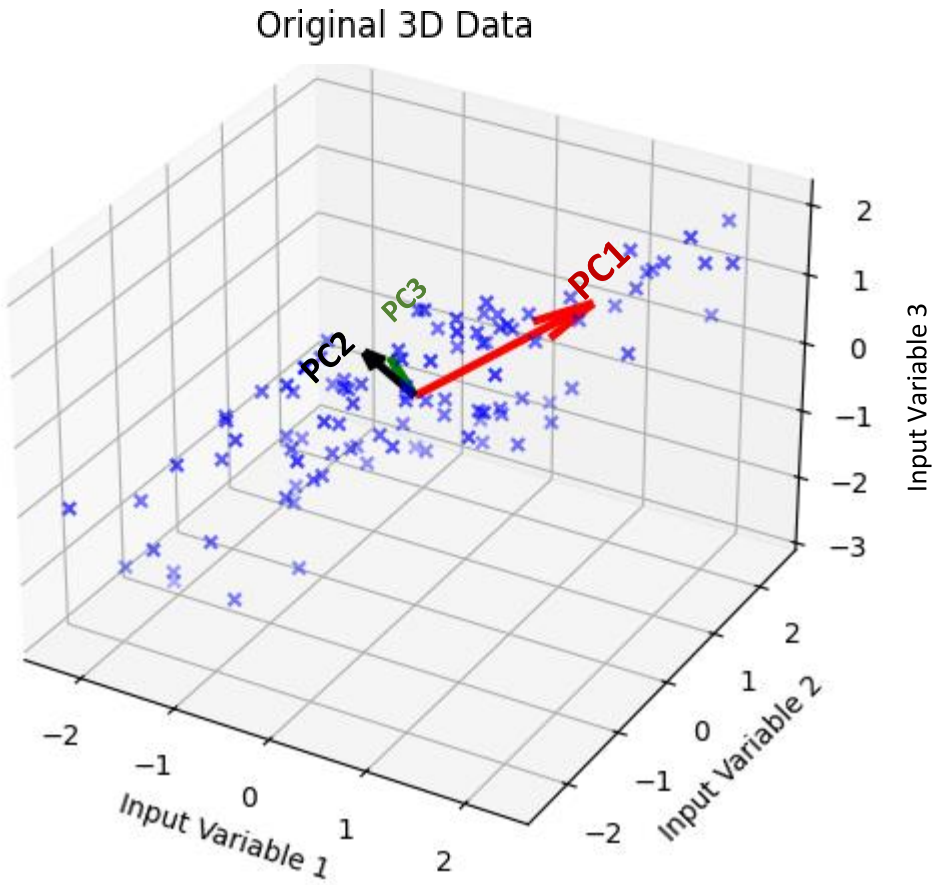
Introduction

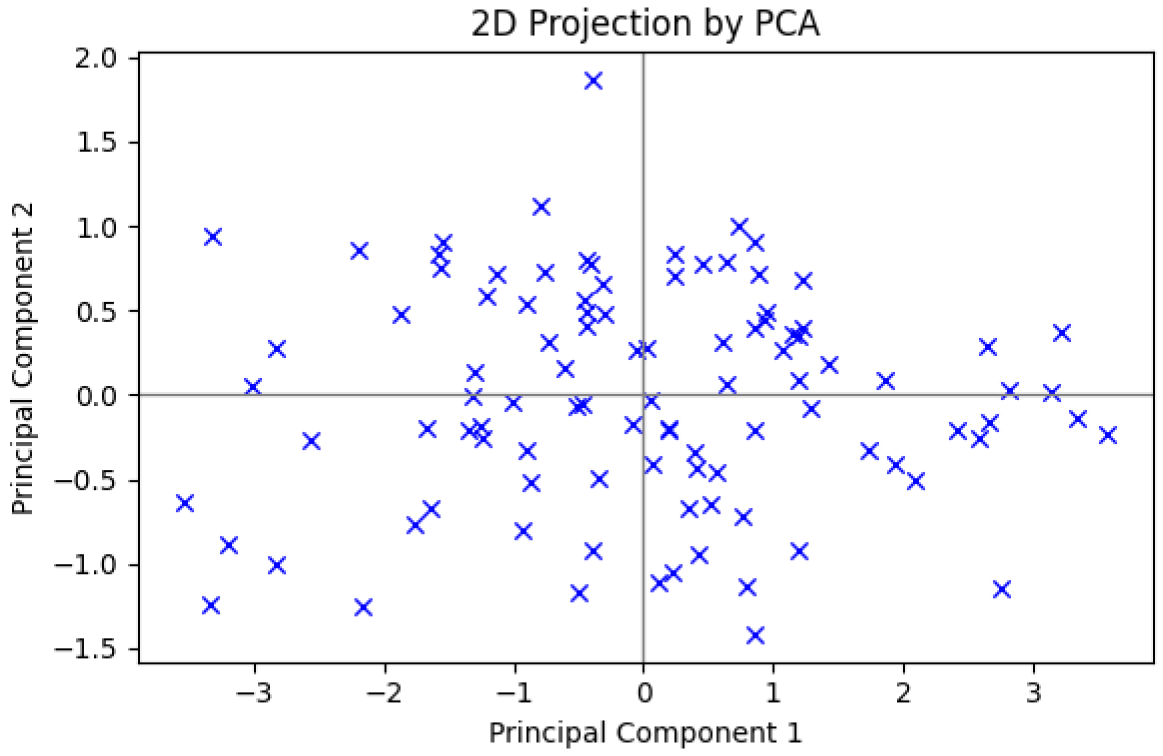
- Principal components analysis (PCA) refers to the process by which principal components are computed, and the subsequent use of these components in understanding the data.
- PCA is an unsupervised approach, since it involves only a set of input variables $x_1, x_2, \ldots, x_m$, without considering the target variable
- Some applications:
  - Reducing data dimensionality
  - Producing new input variables to be used in supervised learning problems (feature extraction)
  - Visualizing and exploring data
  - Clustering
  - Data compression

Principal Components Analysis (PCA)

Introduction

+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors (principal components PC_1, PC_2,….PC_m) are orthogonal to each other and map less and less variance in the data set with increasing order.

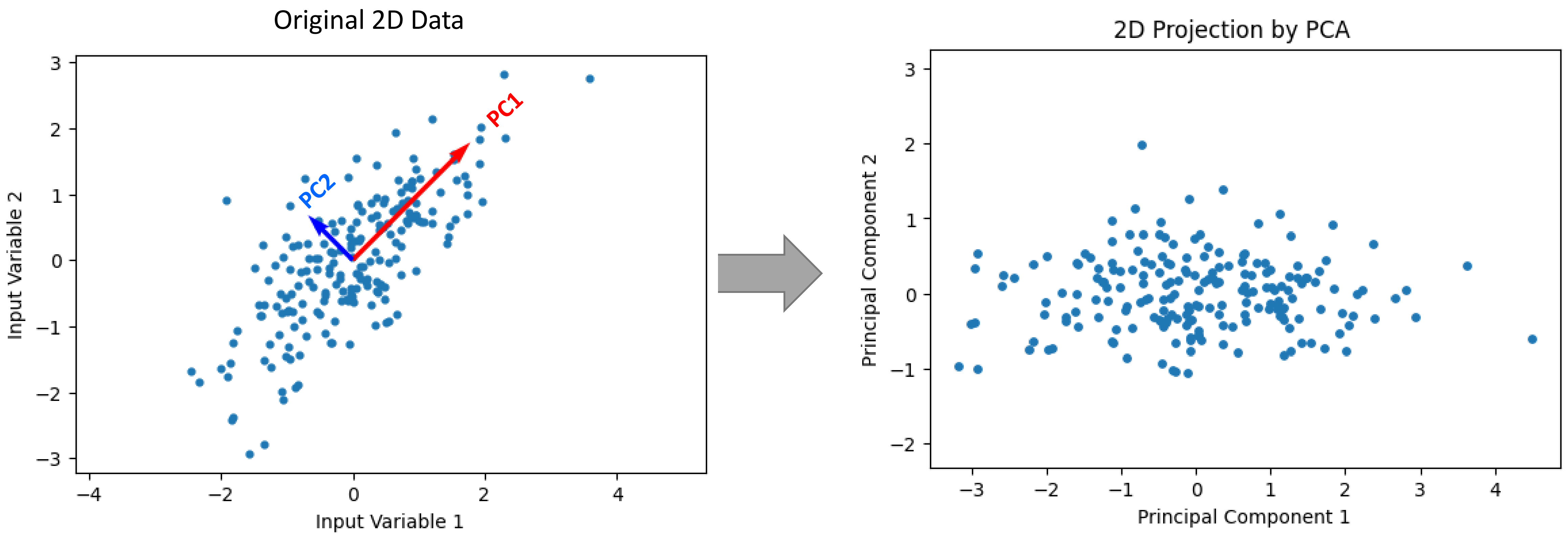




Principal Components Analysis (PCA)

Introduction

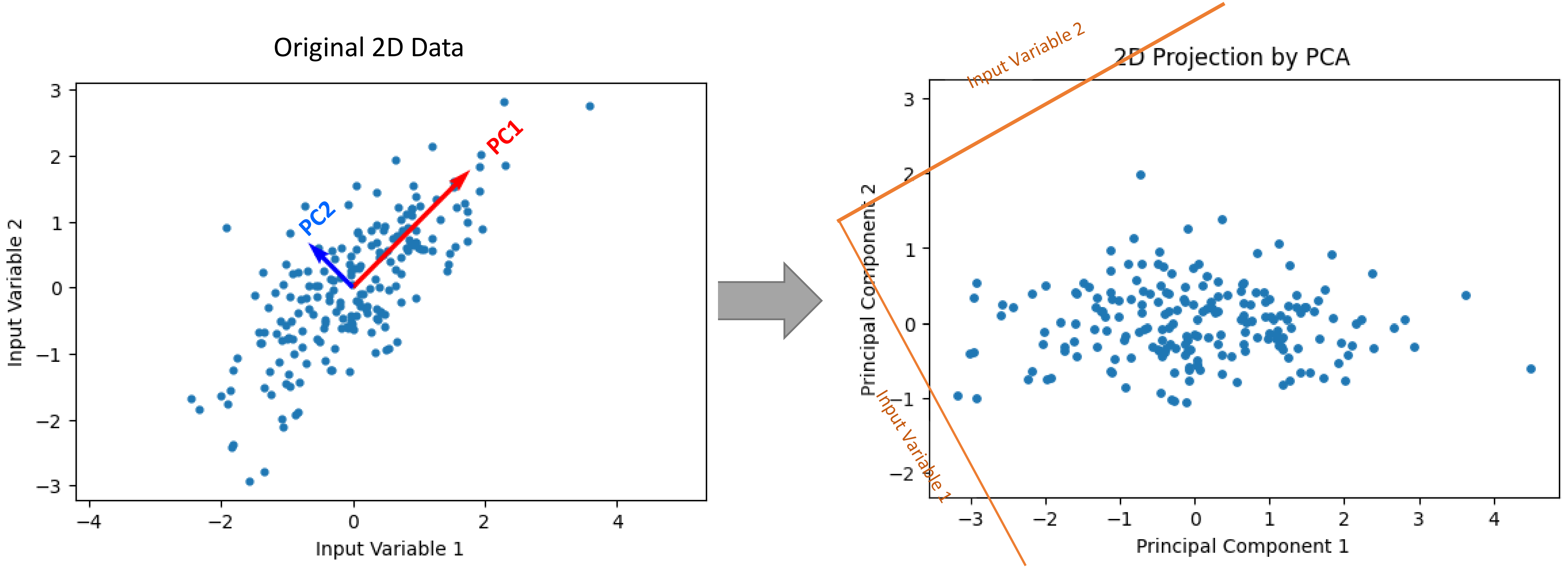
+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors are orthogonal to each other and map less and less variance in the data set with increasing order



Principal Components Analysis (PCA)

Introduction

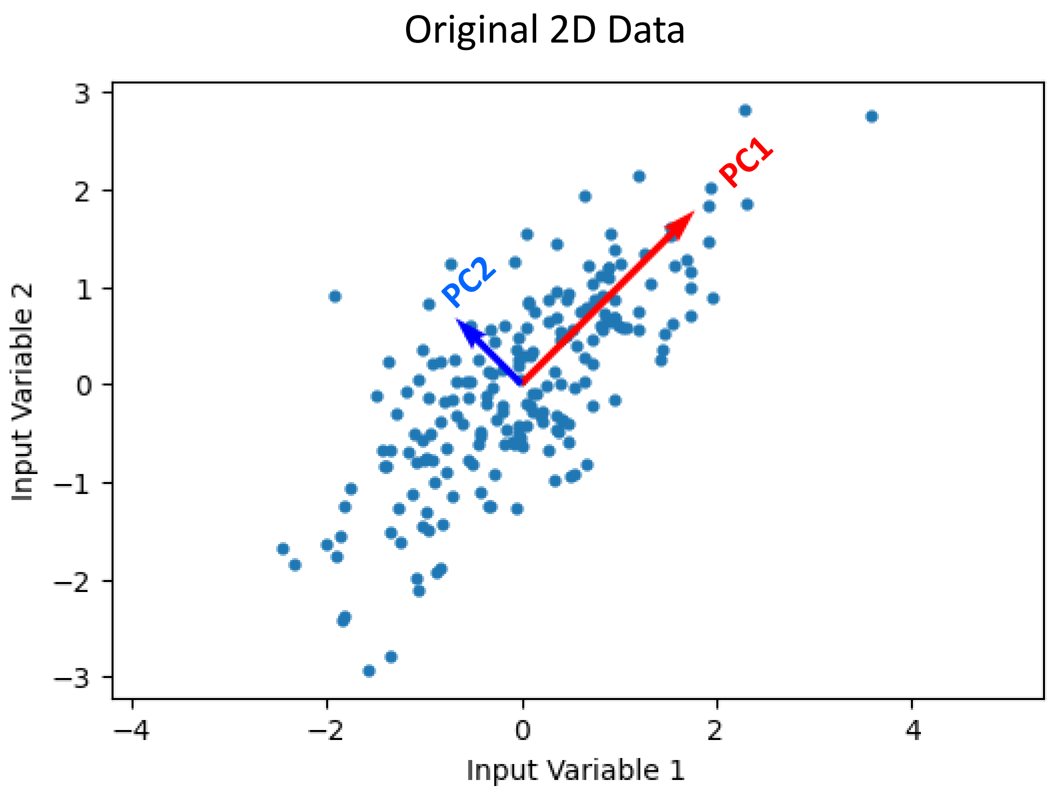
+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors are orthogonal to each other and map less and less variance in the data set with increasing order

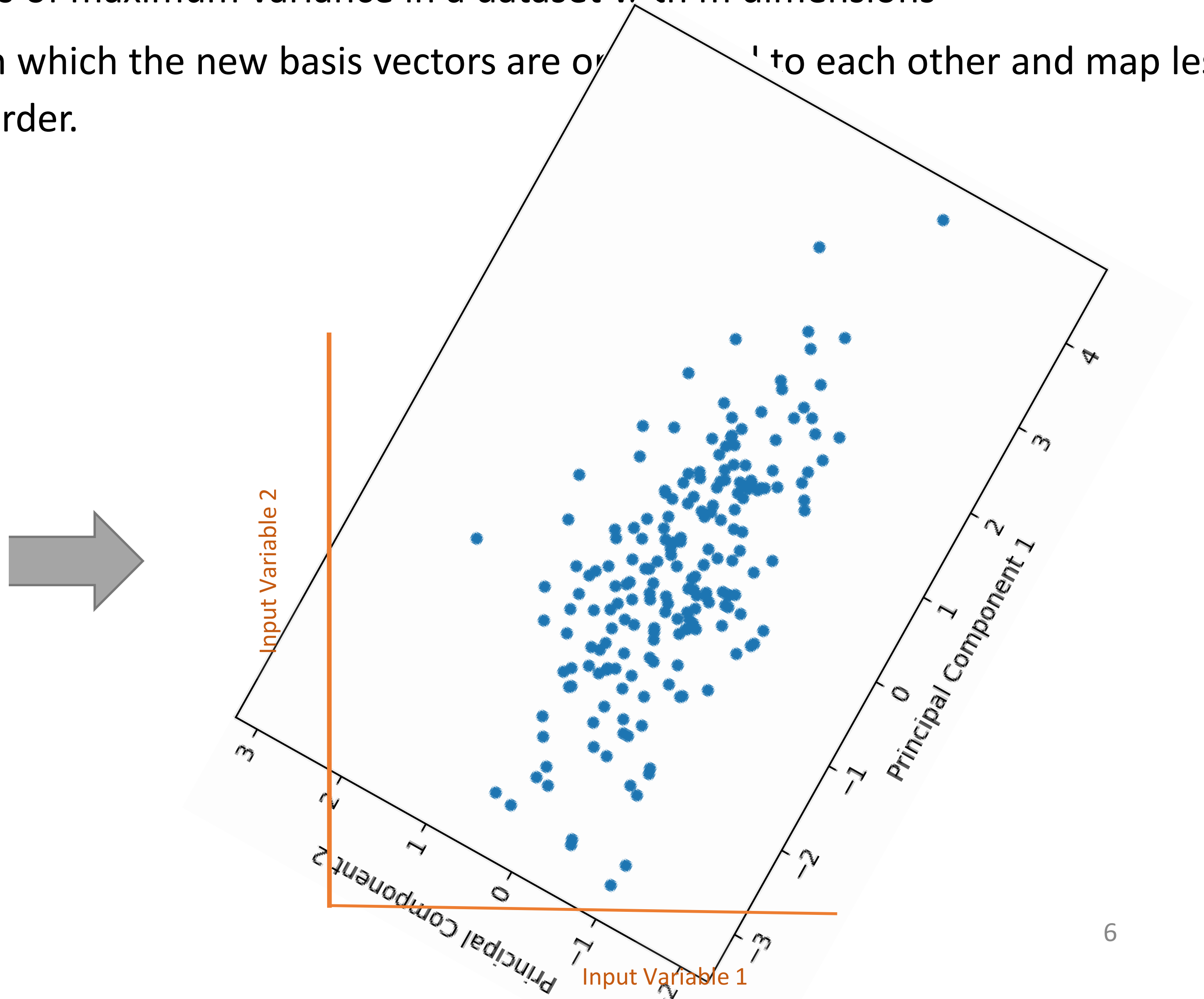


Principal Components Analysis (PCA)

Introduction

+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors are or[illegible]l to each other and map less and less variance in the data set with increasing order.

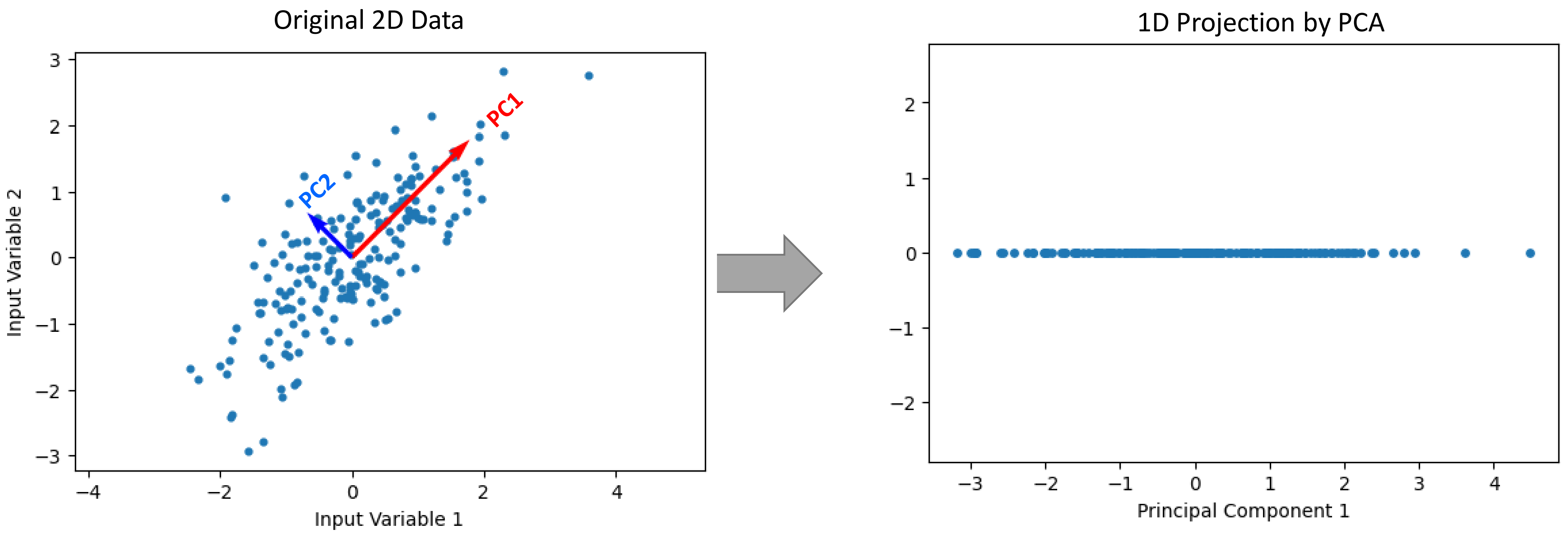




Principal Components Analysis (PCA)

Introduction

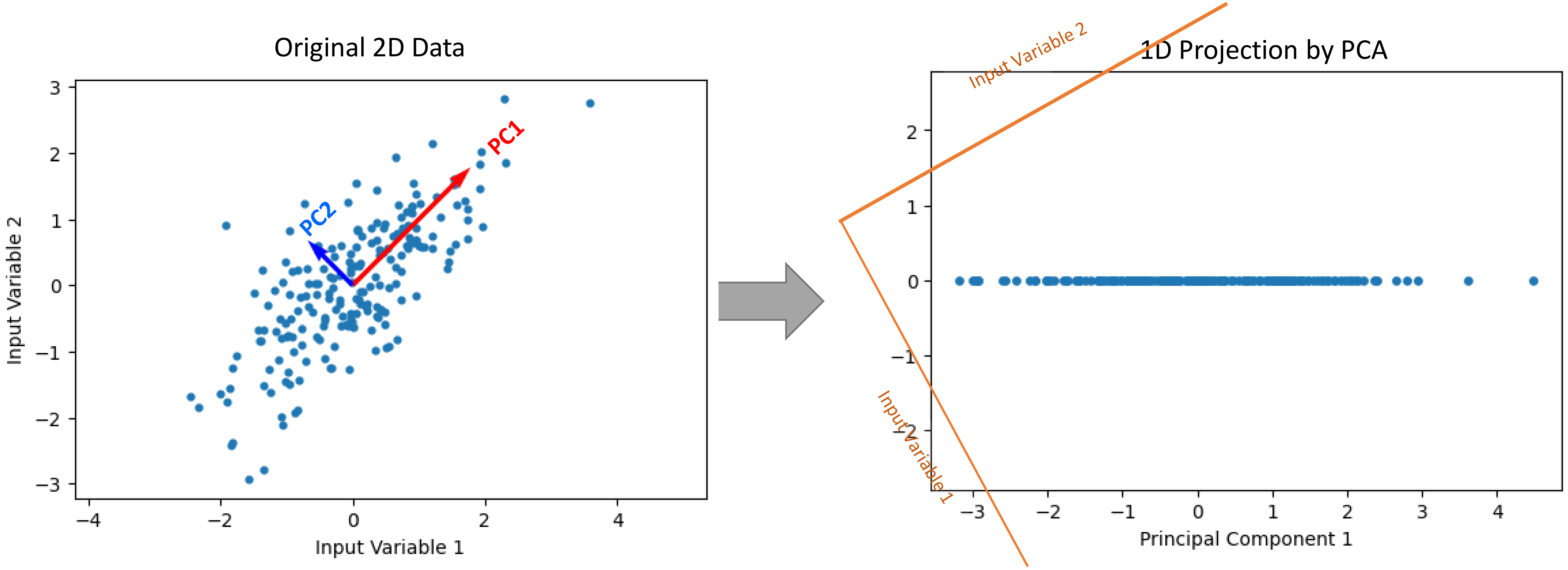
+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors are orthogonal to each other and map less and less variance in the data set with increasing order.



Principal Components Analysis (PCA)

Introduction

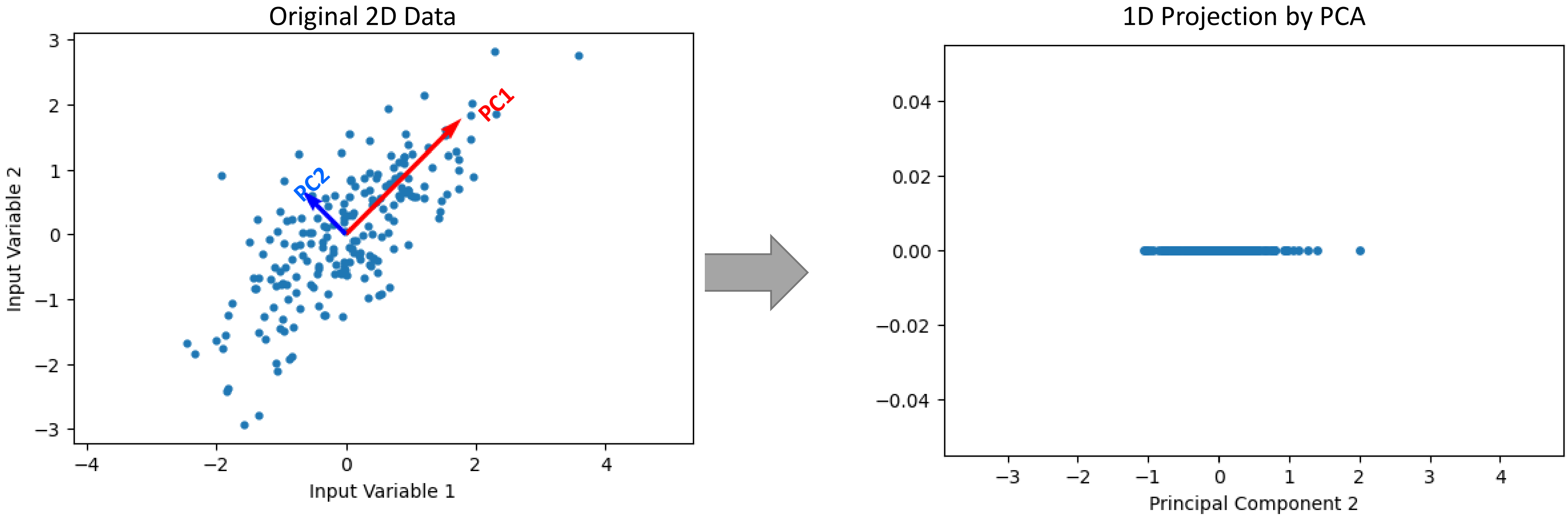
+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors are orthogonal to each other and map less and less variance in the data set with increasing order



Principal Components Analysis (PCA)

Introduction

+ Principal components analysis identifies directions of maximum variance in a dataset with m dimensions
+ Procedure leads to a coordinate transformation in which the new basis vectors are orthogonal to each other and map less and less variance in the data set with increasing order.



Principal Components Analysis (PCA)

Introduction

+ Essentially, we are looking for vectors that they are orthogonal to each other and associated with a value. That value should represent where the data is spread the most.
+ So, we can think about the eigenvectors and the eigenvalues of a matrix. The eigenvectors of a matrix is associated with the eigenvalues
+ But of which matrix? This matrix should be a symmetrical matrix to have orthogonal eigenvectors.
+ Considering the fact that the sum of eigenvalues of a squared Matrix is always equal to its trace (that is, the sum of the diagonal elements)
+ So we search for a matrix that is squared, symmetrical and the sum of its diagonal elements represents the variance of the input variables.
+ What about the matrix of the input variables? But it can be unsquared or asymmetrical
+ What about the covariance matrix of the input variable?

Principal Components Analysis (PCA)

Introduction

- It is useful to have a measure to find out how much the input variables vary from the mean with respect to each other.
- Covariance is such a measure. Covariance is always measured between 2 variables. If you calculate the covariance between one variable and itself, you will get the its variance.
- So, if there is a 3-dimensional dataset x_1, x_2, x_3 then the covariance between the x_1 and x_2 variables, the x_1 and x_3 variables, and the x_2 and x_3 variables could be measured.
- Measuring the covariance between x_1 and x_1 or x_2 and x_2 or x_3 and x_3 would give you the variance of the x_1, x_2 and x_3 variables respectively.

- $Cov(x_1, x_2) = \frac{\sum_{i=1}^{n}(x_{1i}-\overline{X_1})(x_{2i}-\overline{X_2})}{n-1}$

- $Cov(x_1, x_1) = \frac{\sum_{i=1}^{n}(x_{1i}-\overline{X_1})(x_{1i}-\overline{X_1})}{n-1} = \frac{\sum_{i=1}^{n}(x_{1i}-\overline{X_1})^2}{n-1} = \mathrm{Var}(x_1)$

Principal Components Analysis (PCA)

Introduction

+ For a dataset of input variables $x_1, x_2, \ldots, x_m$, without considering the target variable, the covariance matrix of transposed input variables matrix X^T, which has a dimension of (m x n), is as follow:

$$\mathrm{Cov}(X^T) = \begin{bmatrix} Var(x_1) & \ldots & Cov(x_m, x_1) \\ \vdots & & \vdots \\ Cov(x_{1,}, x_m) & \ldots & Var(x_m) \end{bmatrix} = \begin{bmatrix} \frac{\sum_{i=1}^{n}(x_{1i}-\overline{X_1})^2}{n-1} & \ldots & \frac{\sum_{i=1}^{n}(x_{mi}-\overline{X_m})(x_{1i}-\overline{X_1})}{n-1} \\ \vdots & & \vdots \\ \frac{\sum_{i=1}^{n}(x_{1i}-\overline{X_1})(x_{mi}-\overline{X_m})}{n-1} & \ldots & \frac{\sum_{i=1}^{n}(x_{mi}-\overline{X_m})^2}{n-1}) \end{bmatrix}$$

+ For m input variables and n observations, the $\mathrm{Cov}(X^T)$ has a dimension of (m x m)

+ Covariance matrix is a squared and symmetrical matrix as $Cov(x_m, x_1)$ equals $Cov\left(x_{1,}, x_m\right)$

+ If the mean values of the input variables are zero because of the standardization or the centering of the dataset, the expression is simplified to

$$Cov\left(X_{std}^T\right) = \begin{bmatrix} \frac{\sum_{i=1}^{n}(x_{1i})^2}{n-1} & \ldots & \frac{\sum_{i=1}^{n}(x_{mi})(x_{1i})}{n-1} \\ \vdots & & \vdots \\ \frac{\sum_{i=1}^{n}(x_{1i})(x_{mi})}{n-1} & \ldots & \frac{\sum_{i=1}^{n}(x_{mi})^2}{n-1}) \end{bmatrix} = \frac{1}{n-1}\begin{bmatrix} \overset{\text{Var}(x_1)}{\sum_{i=1}^{n}(x_{1i})^2} & \ldots & \sum_{i=1}^{n}(x_{mi})\,(x_{1i}) \\ \vdots & & \vdots \\ \sum_{i=1}^{n}(x_{1i})(x_{mi}) & \ldots & \overset{\text{Var}(x_m)}{\sum_{i=1}^{n}(x_{mi})^2} \end{bmatrix} = \frac{1}{n-1}X_{std}^T X_{std}$$

+ The sum of the diagonal elements is the sum of each input variable variance (is 1 after data standardization) and that represents the total variance in the dataset

Principal Components Analysis (PCA)

Introduction

+ The eigenvectors of this matrix $\frac{1}{n-1}X_{std}^{T}X_{std}$ are unit length vectors, orthogonal to each other and associated with a distinct eigenvalue that represents a ratio of the total variance in the dataset.
+ Therefore, the approaches of computing the principle components depend on calculating the eigenvectors and the eigenvalues of that matrix.
+ The eigenvector, which is associated with the highest eigenvalue, is the first principle component. The second principle component is associated with the second highest eigenvalue and so on.
+ Principal Components Analysis (PCA) can be conducted through two primary mathematical approaches:
    - Using the eigendecomposition of the covariance matrix of the standardized or centered data
    - Applying Singular Value Decomposition (SVD) directly to the standardized or centered data matrix itself.
+ Both methods will yield the same principal components and can be used to perform PCA, but they differ in their computational properties and ease of use.

Principal Components Analysis (PCA)

1- Eigendecomposition of the Covariance Matrix

+ This approach can be conducted by the following steps:

1. Standardize the data: each input variable with zero mean and one as standard deviation
2. Calculate the covariance matrix of the standardized data
3. Calculate eigenvectors and eigenvalues of that covariance matrix
4. Order the eigenvectors with respect to their eigenvalues λ, highest to lowest (in descending order)

   $\vec{v}_1, \vec{v}_2, \ldots\ldots\ldots, \vec{v}_m$, where $\lambda_1 > \lambda_2 > \ldots\ldots\ldots . \lambda_m$
5. Construct a projection matrix (you can choose a number of principle components to reduce the original dimension of the data. Assuming the new dimension is p, the projection matrix will be as follows:

   $$\text{Projection Matrix} = V = (\vec{v}_1, \vec{v}_2, \ldots\ldots\ldots, \vec{v}_p)$$
6. Data transformation

   $$\text{Transformed data} = X_{std}.V \quad ,$$

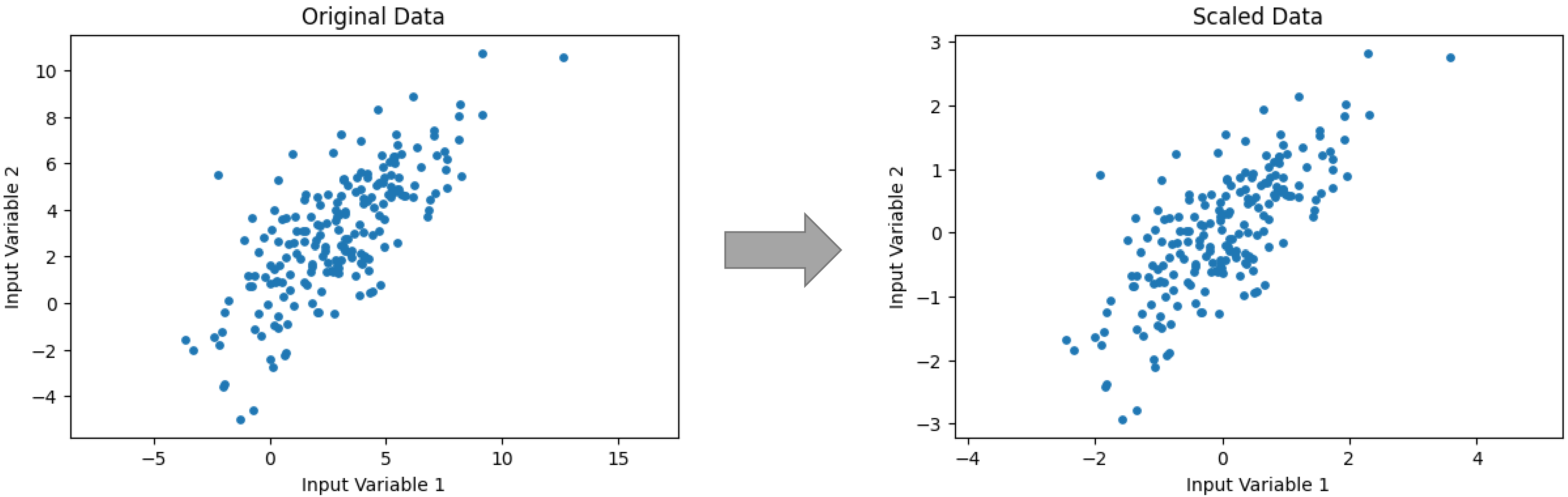
where X_{std} has a dimension of (n x m), V has a dimension of (m x p) and the transformed data has a dimension of (n x p)

Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

+ This approach can be conducted by the following steps:

    1. Standardize the data:



Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

+ This approach can be conducted by the following steps:

    1. Standardize the data:

    2. Calculate the covariance matrix:

$$Cov\left(X_{std}^{T}\right) = \begin{bmatrix} 1 & 0.7469 \\ 0.7469 & 1 \end{bmatrix}$$

Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix — Example

+ This approach can be conducted by the following steps:

1. Standardize the data.
2. Calculate the covariance matrix
3. Calculate eigenvectors and eigenvalues

- Eigenvalues result from the equation

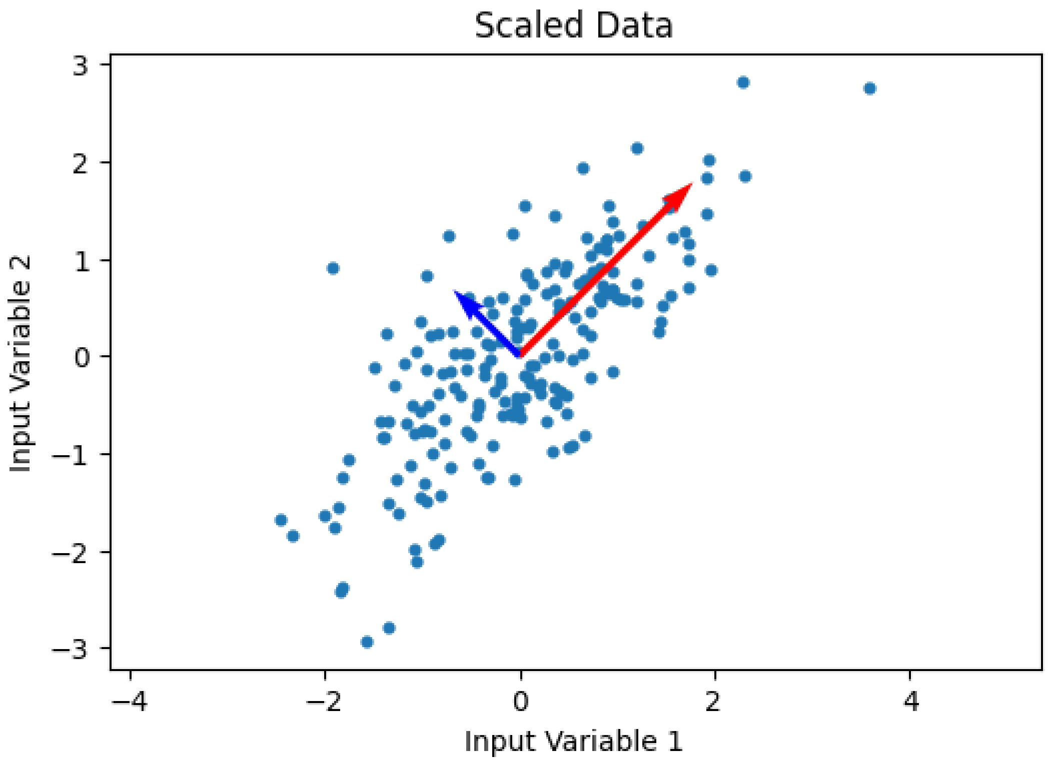
$$\det\left(\mathbf{X}^T \cdot \mathbf{X} - \lambda \cdot \mathrm{I}\right) = 0$$

```
eigen_vals = [1.7469 0.2531]
```

- For every eigenvalue λ_i there is an eigenvector, it is calculated by the equation

$$(\mathbf{X}^T \cdot \mathbf{X} - \lambda_i \cdot \mathrm{I})\ \vec{v_i} = 0$$

$$\texttt{eigen\_vecs} = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}$$



Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

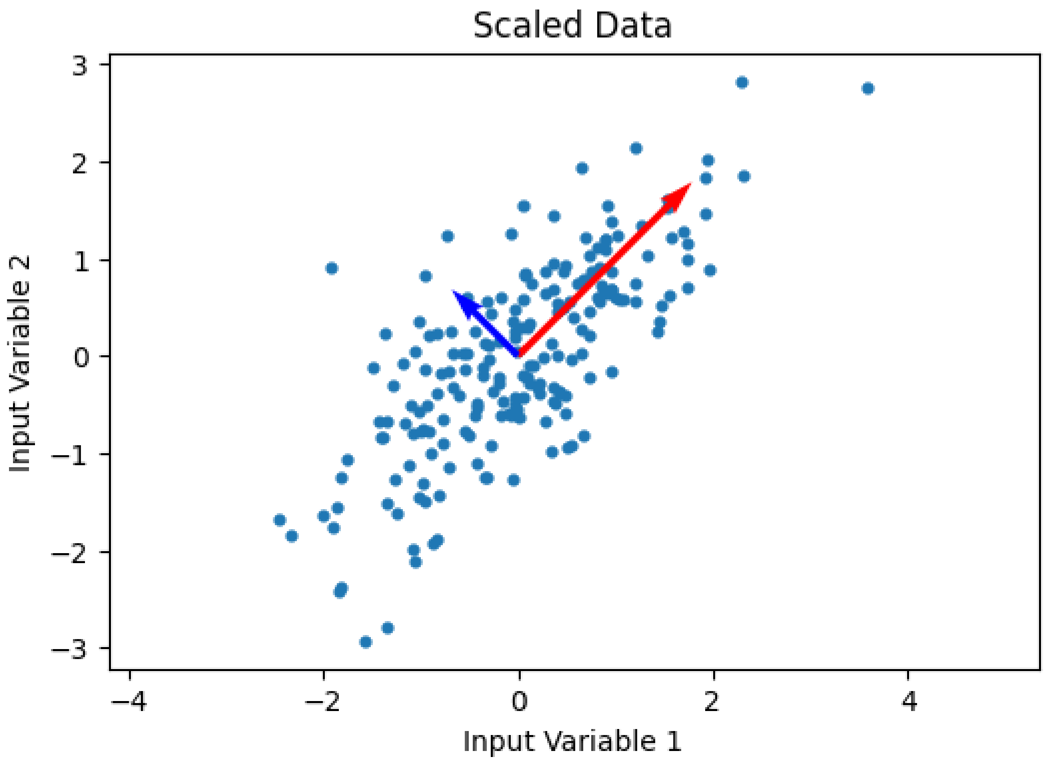
+ This approach can be conducted by the following steps:

1. Standardize the data: each input variable with zero mean and one as standard deviation
2. Calculate the covariance matrix: calculate the covariance matrix of the standardized data.
3. Calculate eigenvectors and eigenvalues of the covariance matrix
4. Order the eigenvectors with respect to their eigenvalues λ, highest to lowest

$$\vec{v}_1, \vec{v}_2, \ldots\ldots, \vec{v}_m, \text{ where } \lambda_1 > \lambda_2 > \ldots\ldots\ldots\lambda_m$$

$\vec{v}_1 = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix}$ is associated with $\lambda_1 = 1.7469$

$\vec{v}_2 = \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix}$ is associated with $\lambda_2 = 0.2531$



Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

+ This approach can be conducted by the following steps:

1. Standardize the data: each input variable with zero mean and one as standard deviation
2. Calculate the covariance matrix: calculate the covariance matrix of the standardized data.
3. Calculate eigenvectors and eigenvalues of the covariance matrix
4. Order the eigenvectors with respect to their eigenvalues λ, highest to lowest

   $\vec{v}_1, \vec{v}_2, \ldots\ldots\ldots, \vec{v}_m$, where $\lambda_1 > \lambda_2 > \ldots \ldots \ldots . \lambda_m$

5. Construct a projection matrix (you can choose a number of principle components to reduce the original dimension of the data. Assuming the new dimension is p, the projection matrix will be as follows:

$$\text{Projection Matrix} = V = (\vec{v}_1, \vec{v}_2, \ldots\ldots\ldots, \vec{v}_p)$$

$$V = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}, \text{ as } p=2$$

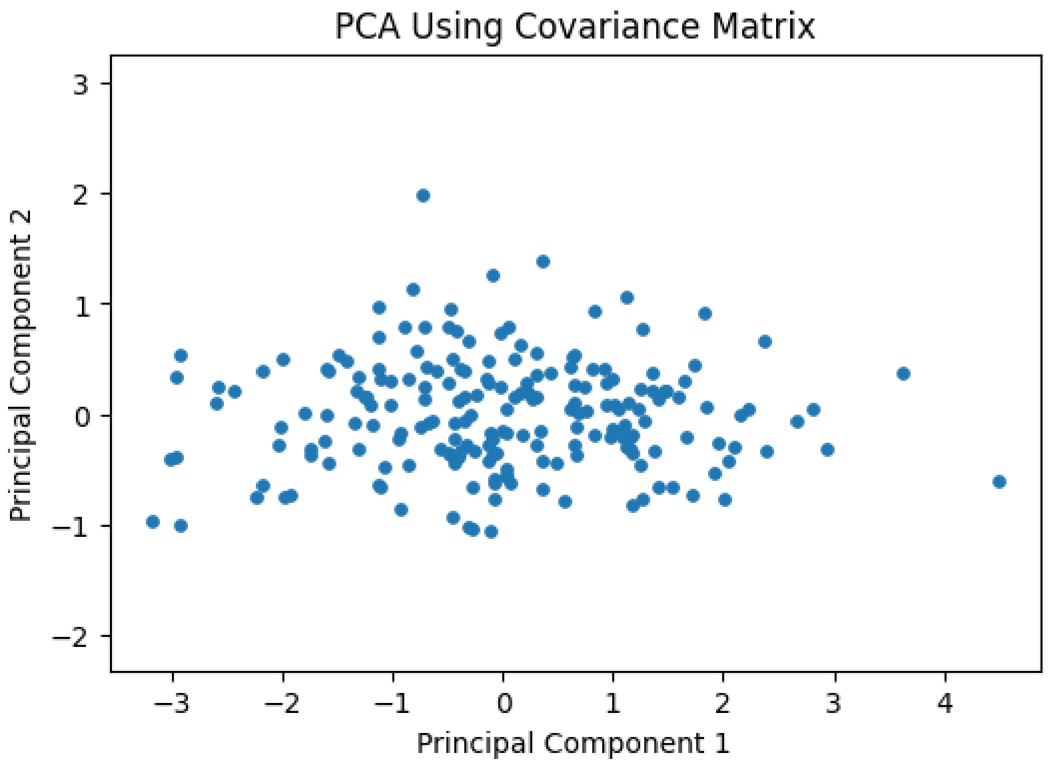
Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

+ This approach can be conducted by the following steps:

1. Standardize the data
2. Calculate the covariance matrix
3. Calculate eigenvectors and eigenvalues of the covariance matrix
4. Order the eigenvectors with respect to their eigenvalues λ, highest to lowest
5. Construct a projection matrix (you can choose a number of principle components to reduce the original dimension of the data.
6. Data transformation

Transformed data= X_{std}.V , where X_{std} has a dimension of (200 x 2), V has a dimension of (2 x 2) and the transformed data has a dimension of (200 x 2)



Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

+ This approach can be conducted by the following steps:

1. Standardize the data: each input variable with zero mean and one as standard deviation
2. Calculate the covariance matrix: calculate the covariance matrix of the standardized data.
3. Calculate eigenvectors and eigenvalues of the covariance matrix
4. Order the eigenvectors with respect to their eigenvalues λ, highest to lowest

   $\vec{v}_1, \vec{v}_2, \ldots\ldots\ldots, \vec{v}_m$, where $\lambda_1 > \lambda_2 > \ldots \ldots \ldots . \lambda_m$

5. Construct a projection matrix (you can choose a number of principle components to reduce the original dimension of the data. Assuming the new dimension is p, the projection matrix will be as follows:

$$\text{Projection Matrix} = V = (\vec{v}_1, \vec{v}_2, \ldots\ldots\ldots, \vec{v}_p)$$

$$V = \begin{bmatrix} 0.7071 \\ 0.7071 \end{bmatrix}, \text{ as } p=1$$

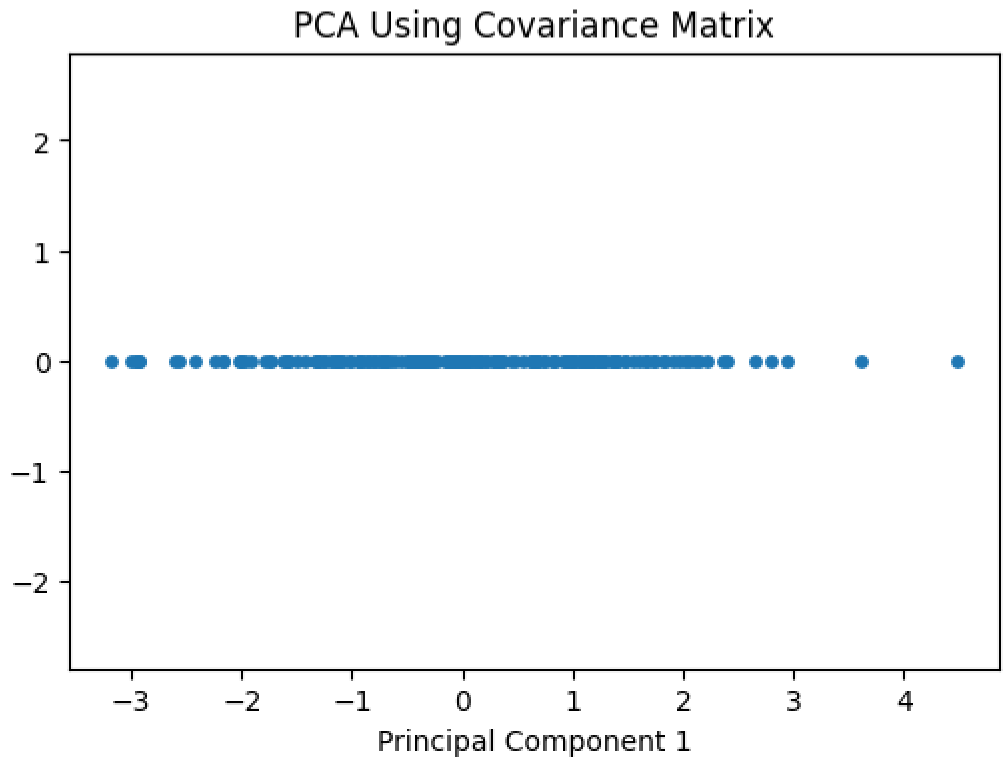
Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – Example

+ This approach can be conducted by the following steps:

1. Standardize the data
2. Calculate the covariance matrix
3. Calculate eigenvectors and eigenvalues of the covariance matrix
4. Order the eigenvectors with respect to their eigenvalues λ, highest to lowest
5. Construct a feature vector (you can choose a number of principle components to reduce the original dimension of the data.
6. Data transformation

Transformed data= X_{std}.V , where X_{std} has a dimension of (200 x 1), V has a dimension of (2 x 1) and the transformed data has dimension of (200 x 1)

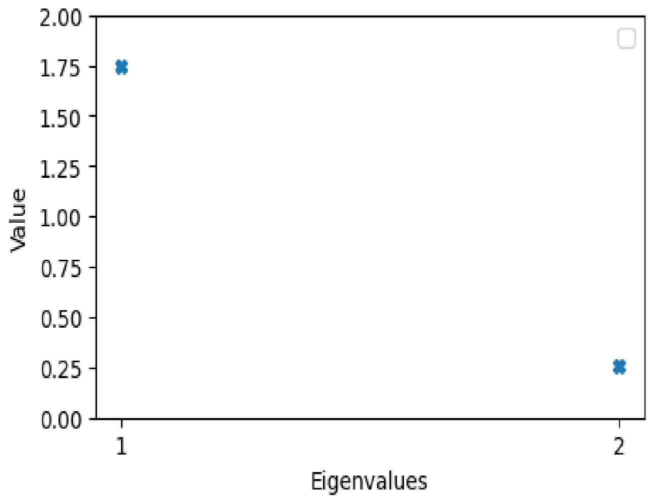


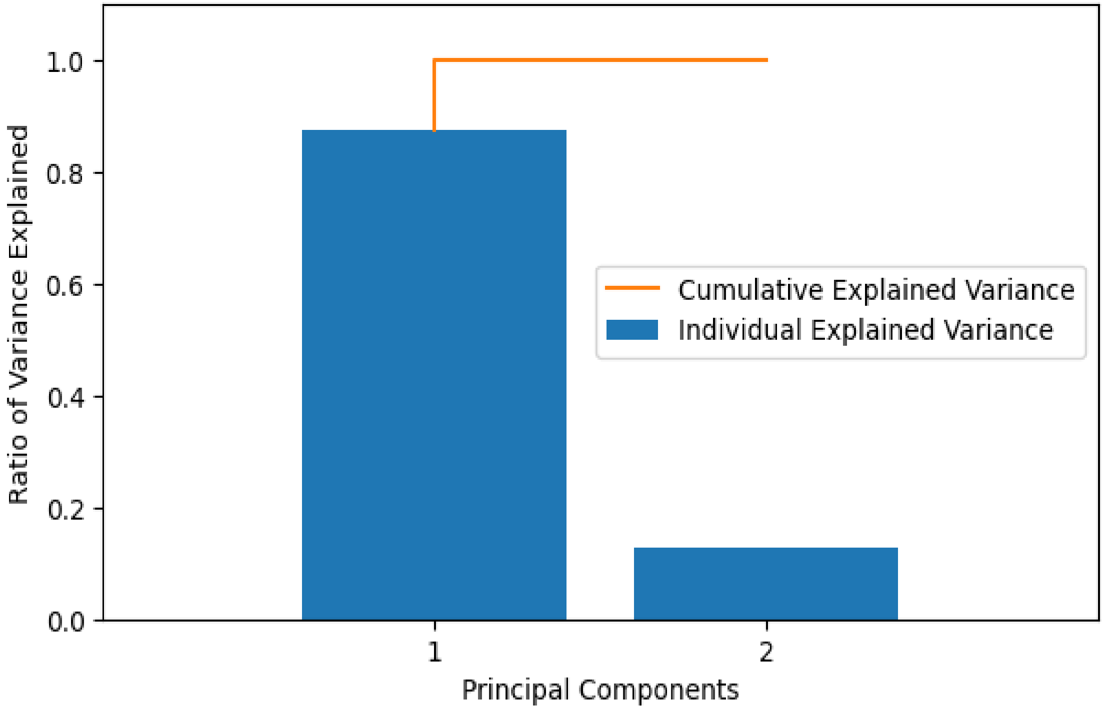
Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix — Example

Explained Variance:

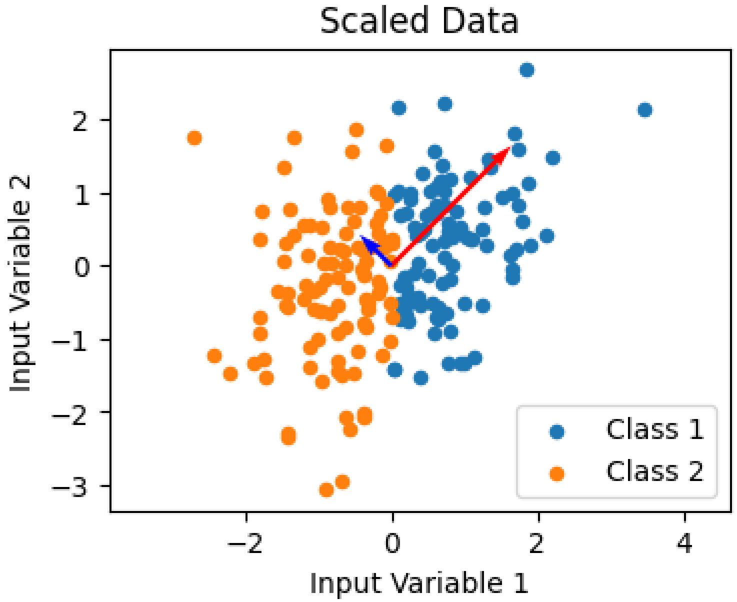
+ How many principal components should you keep for the new feature space?
+ A useful measure is the so-called "explained variance", which can be calculated from the eigenvalues, indicating how much information (variance) can be assigned to each principal component.
+ Variance that is explained by $\lambda_1 = \frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_1}{\lambda_1 + \lambda_2} \simeq 0.87$
+ Variance that is explained by $\lambda_2 = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_2}{\lambda_1 + \lambda_2} \simeq 0.13$
+ Principle component 1 can represent 87% of the total variance
+ Principle component 2 can represent 13% of the total variance

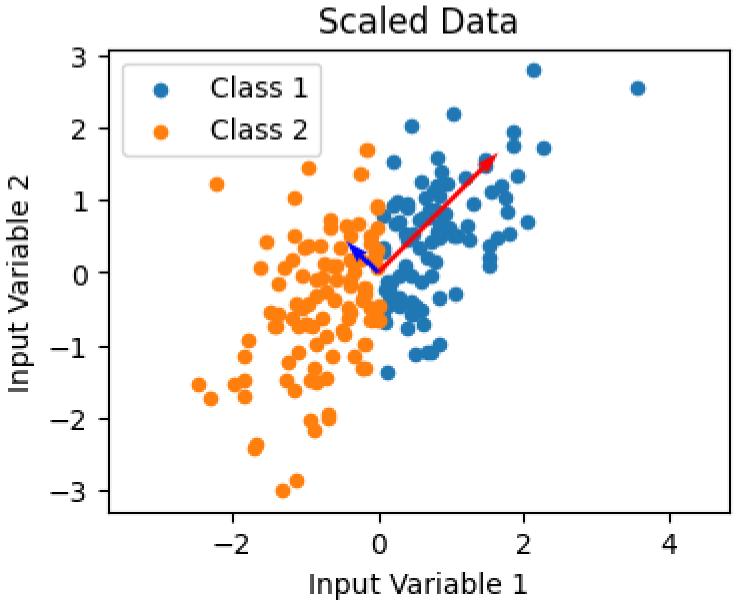


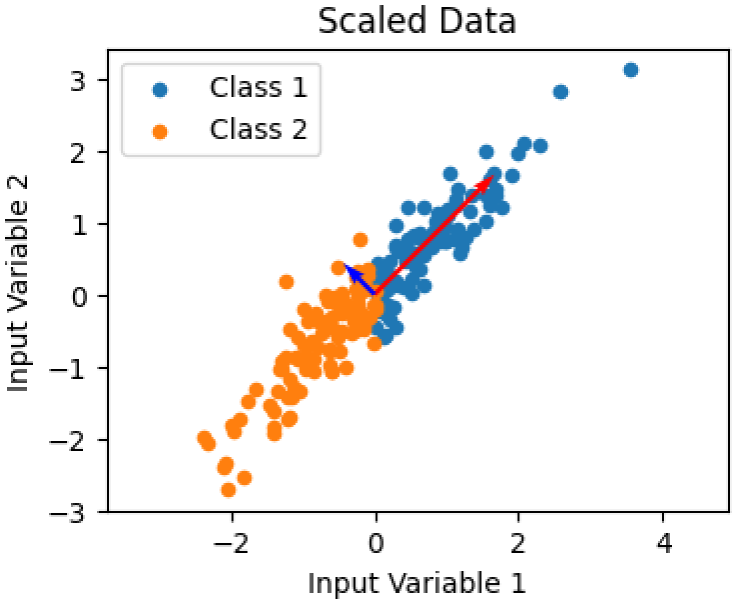


Principal Components Analysis (PCA)

Correlation Between Original Input Variables and λ of Covariance Matrix







$$Cov(X_{std}) = \begin{bmatrix} 1 & 0.3687 \\ 0.3687 & 1 \end{bmatrix}$$

`eigen_values = [1.3687 0.6313]`

$$Cov(X_{std}) = \begin{bmatrix} 1 & 0.6206 \\ 0.6206 & 1 \end{bmatrix}$$

`eigen_values = [1.6206 0.3794]`

$$Cov(X_{std}) = \begin{bmatrix} 1 & 0.9366 \\ 0.9366 & 1 \end{bmatrix}$$

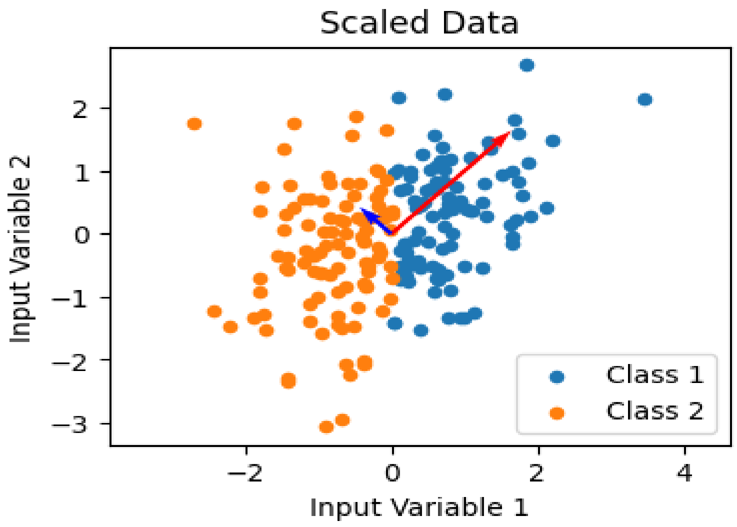
`eigen_values = [1.9366  0.0634]`

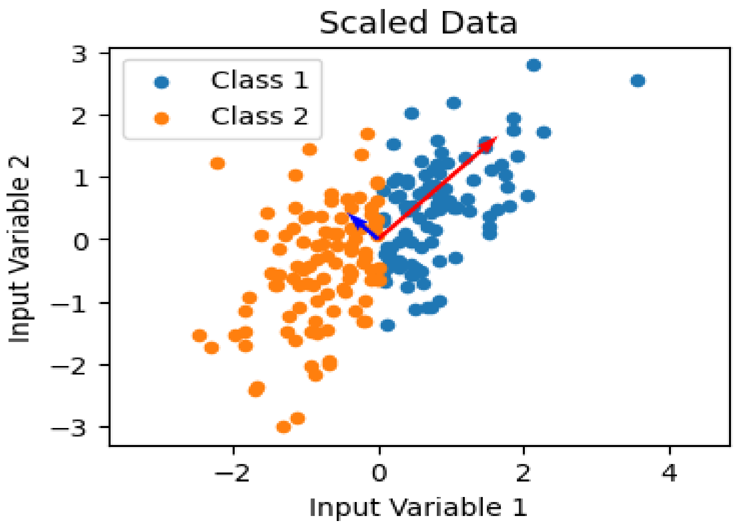
Variance explained by $\lambda_1 = \frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i} \simeq 0.6844$ | $\simeq 0.8103$ | $\simeq 0.9683$

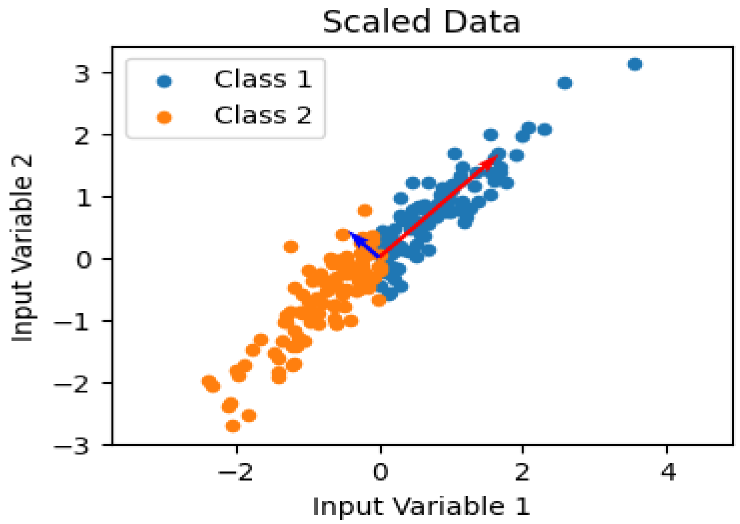
Variance explained by $\lambda_2 = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} \simeq 0.3165$ | $\simeq 0.1897$ | $\simeq 0.0317$

Principal Components Analysis (PCA)

Correlation Between Original Input Variables and λ of Covariance Matrix

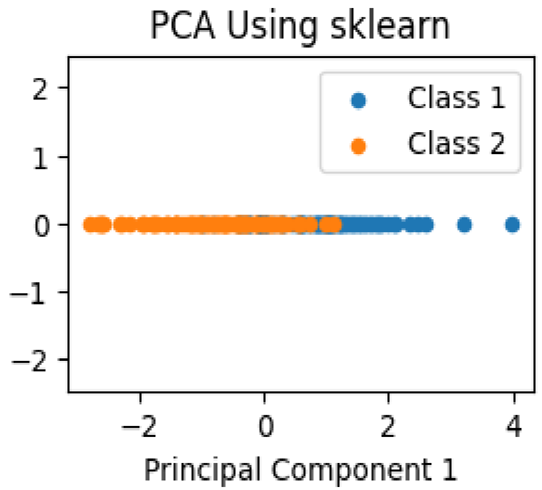


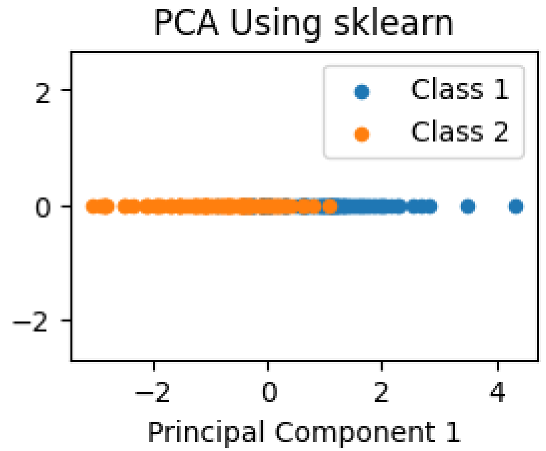


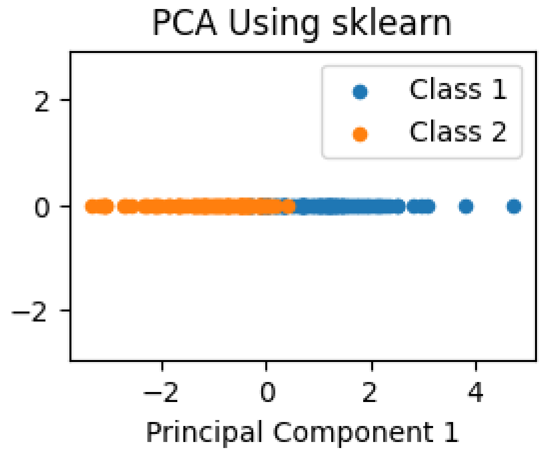


Variance explained by $\lambda_1 = \frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i} \simeq 0.6844$ | $\simeq 0.8103$ | $\simeq 0.9683$

Variance explained by $\lambda_2 = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} \simeq 0.3165$ | $\simeq 0.1897$ | $\simeq 0.0317$





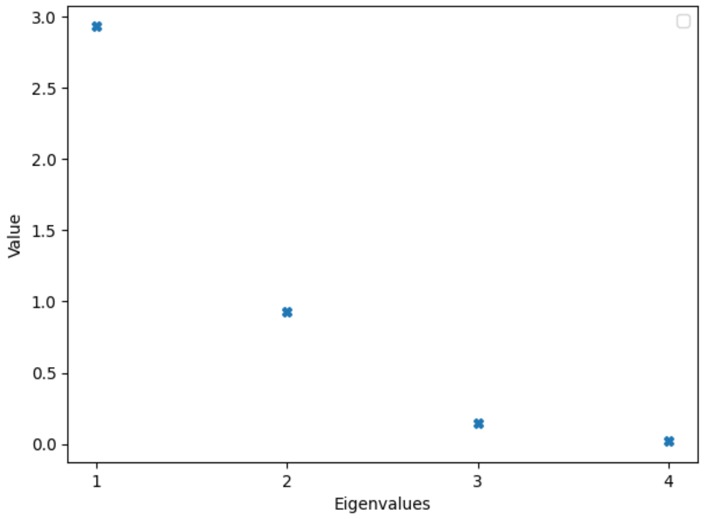


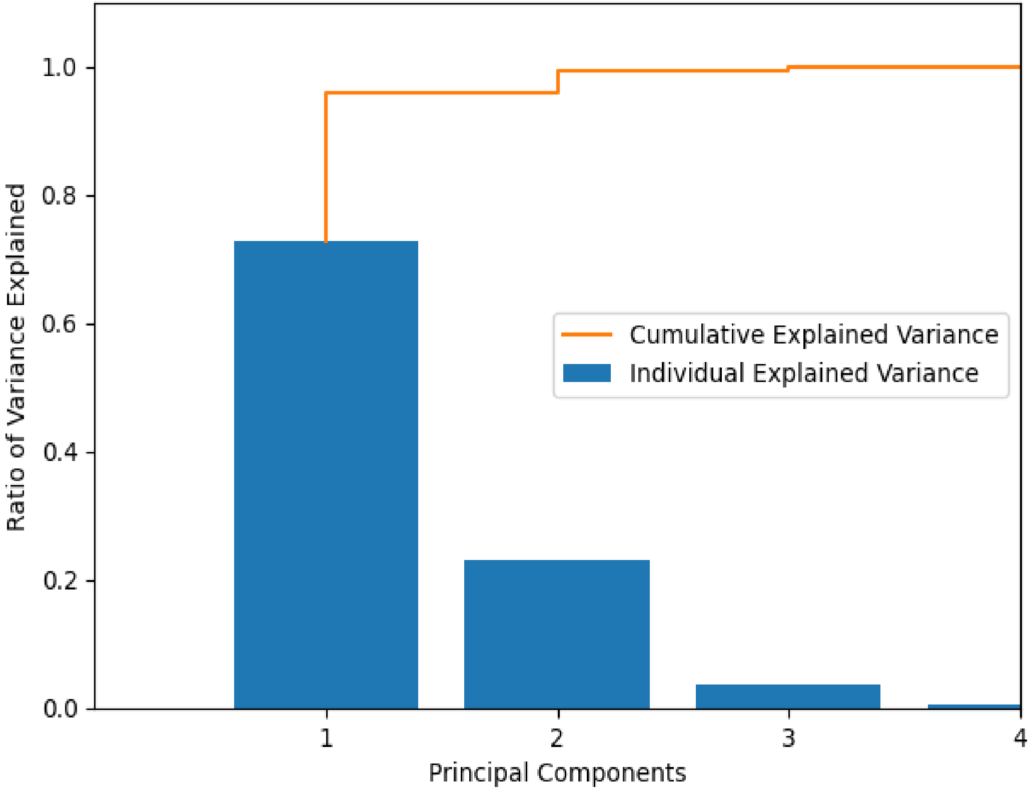
Principal Components Analysis (PCA)

Eigendecomposition of the Covariance Matrix – IRIS-Dataset

Explained Variance:

- How many principal components should you keep for the new feature space?
- Variance that is explained by $\lambda_1 = \frac{\lambda_1}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \simeq 72.77$
- Variance that is explained by $\lambda_2 = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \simeq 23.03$
- Variance that is explained by $\lambda_3 = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \simeq 3.68$
- Variance that is explained by $\lambda_4 = \frac{\lambda_2}{\sum_{i=1}^{m} \lambda_i} = \frac{\lambda_4}{\lambda_1+\lambda_2+\lambda_3+\lambda_4} \simeq 0.52$
- The most (72.77 %) of the variance can be explained by the first principal component alone
- The second principal component still contains some information (23.03 %), while the third and fourth principal components can be safely dropped without losing too much information.





Hochschule Karlsruhe
University of
Applied Sciences

Fakultät für
**Elektro- und
Informationstechnik**

www.h-ka.de