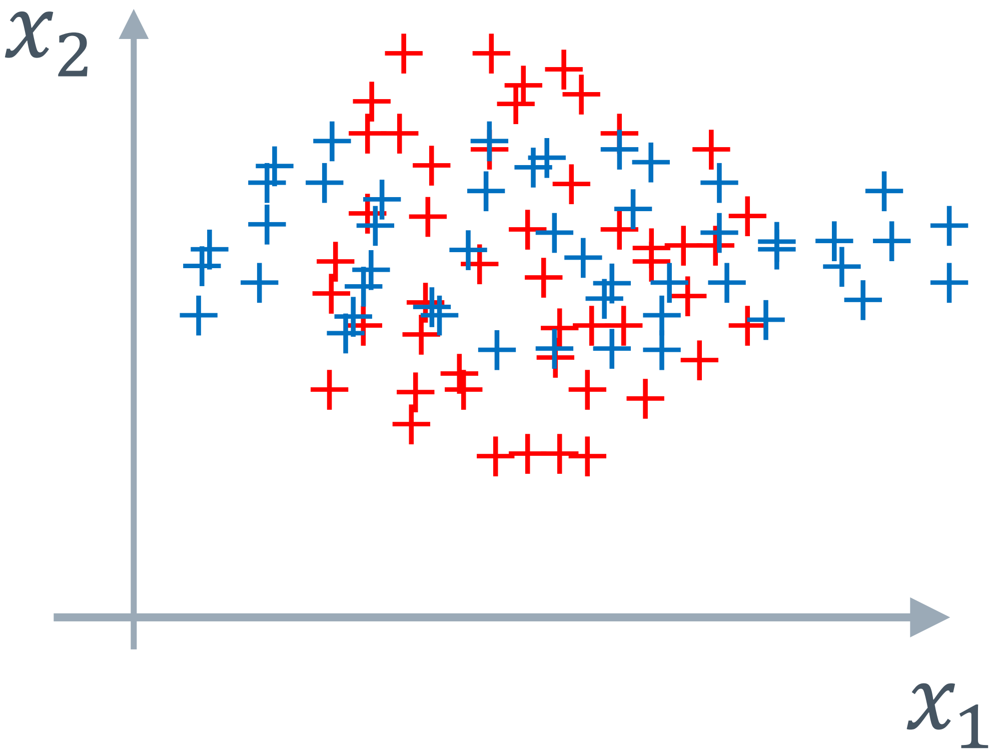
Principal Component Analysis

Contents

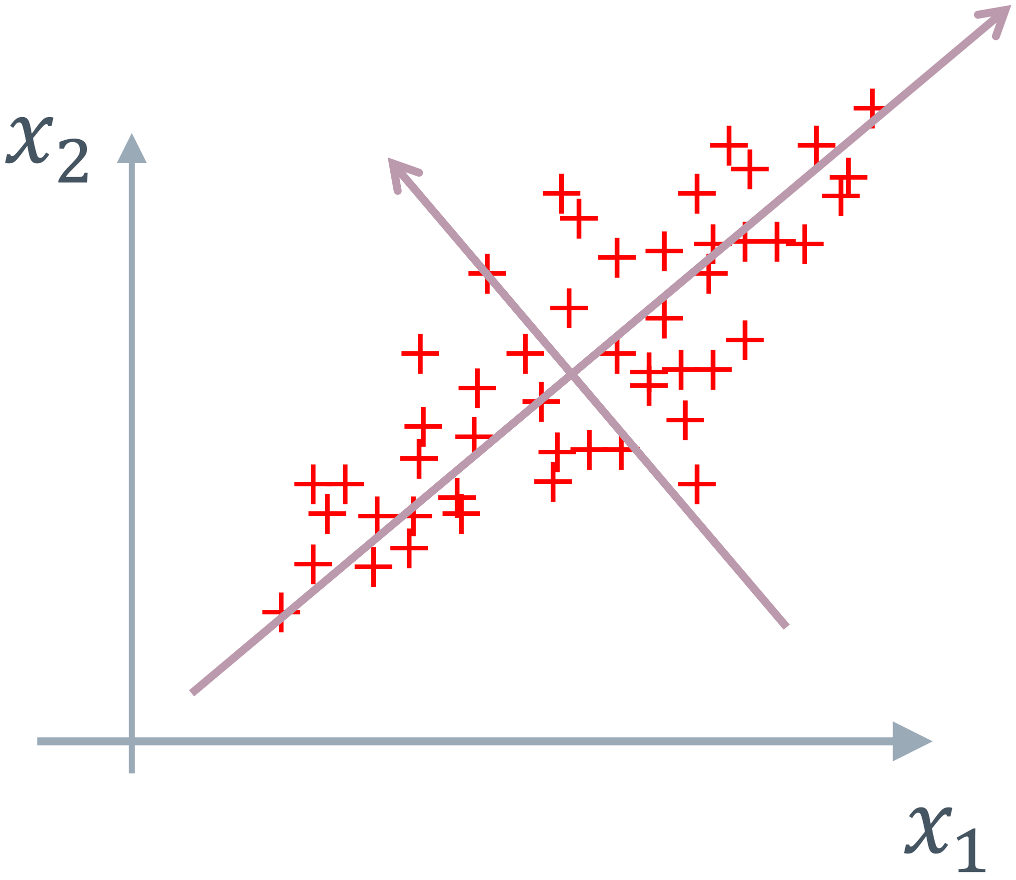
Dr.- Ing. Maggie Mashaly

Importance of PCA

Correlation



Uncorrelated



Correlated

Main Questions

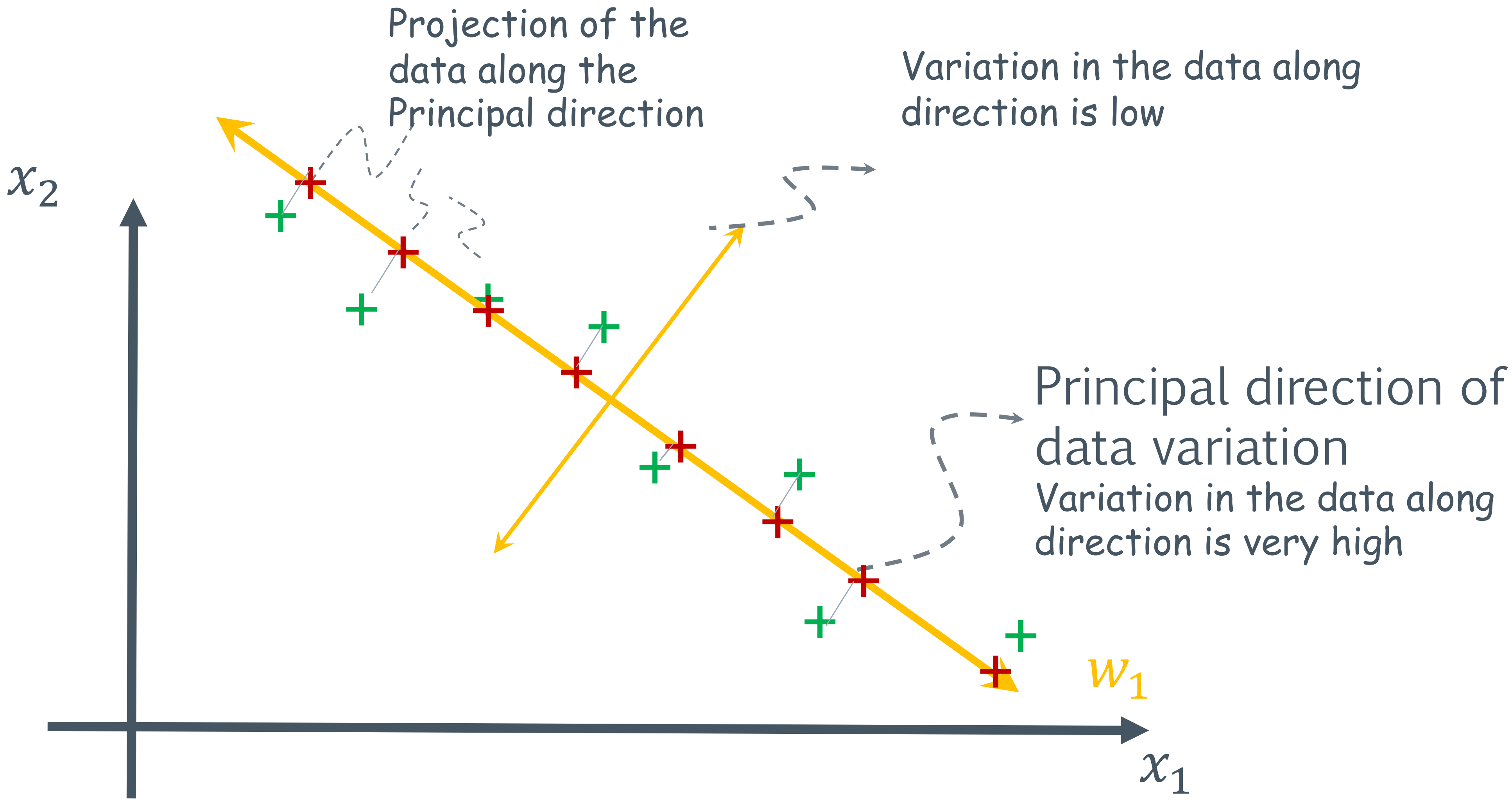
- How to remove correlation?
- What are the dimensions with the most information?
- How can we provide a better scaling or normalization?

Principal Component Analysis

To identify the principal directions in which the data varies

- ❑Better data representation
  - ➢Reduce the number of features
  - ➢Reduce the relation between features
- ❑Eliminate non principal data
- ❑Can be used for compression applications

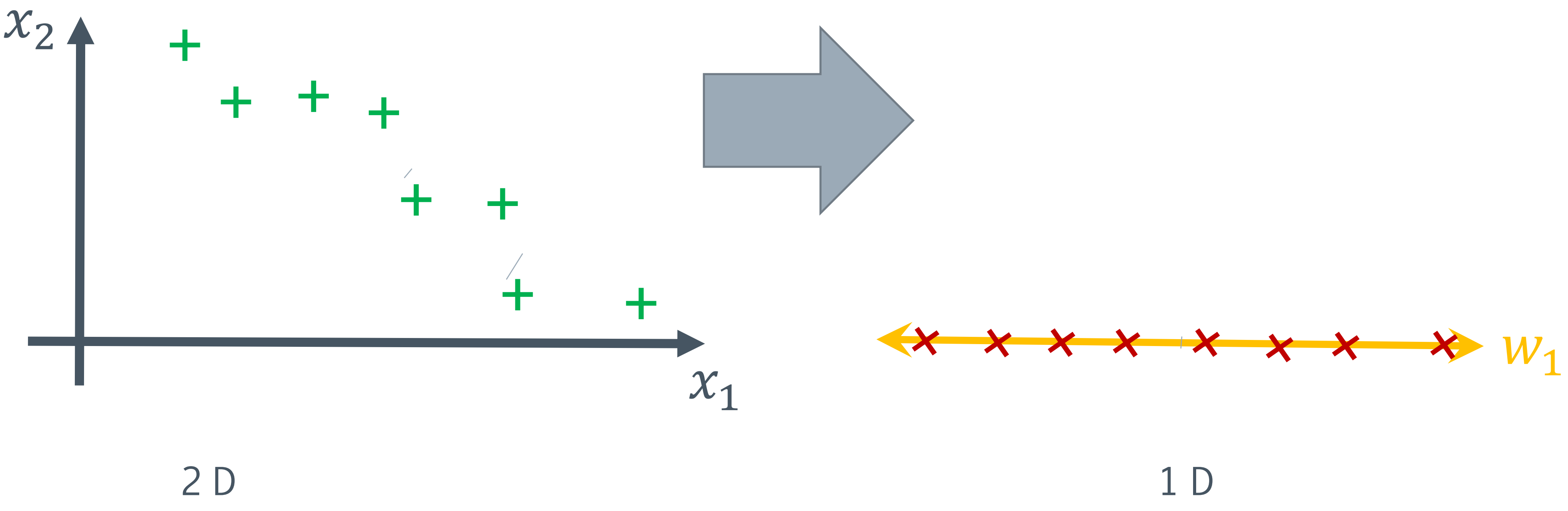
Principal Component Analysis
2D Example



Dimensionality Reduction

Dimensionality Reduction

We are interested to find a linear transformation from an m dimension to n dimension (reduce the number of features from m to n



Dr.- Ing. Maggie Mashaly

Linear Transformation

$$x = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & x_3^{(1)} & \dots x_m^{(1)} \\ x_1^{(2)} & x_2^{(2)} & x_3^{(2)} & \dots x_m^{(2)} \\ x_1^{(3)} & x_2^{(3)} & x_3^{(3)} & \dots x_m^{(3)} \\ & \dots\dots\dots\dots\dots. & & \\ x_1^{(n)} & x_2^{(n)} & x_3^{(n)} & \dots x_m^{(n)} \end{bmatrix}$$

Data Matrix

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & w_{1,3} \dots w_{1,k} \\ w_{2,1} & w_{2,2} & w_{2,3} \dots w_{2,k} \\ w_{3,1} & w_{3,2} & w_{3,3} \dots w_{3,k} \\ & \dots\dots\dots\dots\dots. & \\ w_{m,1} & w_{m,2} & w_{m,2} \dots w_{m,k} \end{bmatrix}$$

Transformation matrix

Transform into nxk

nxk nxm mxk

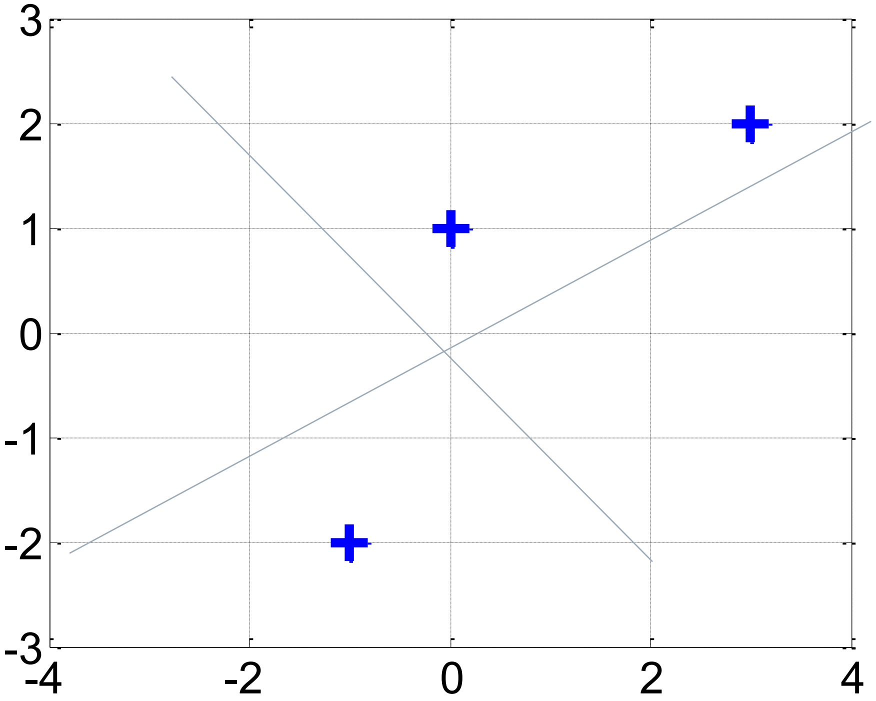
$$Z = xW$$

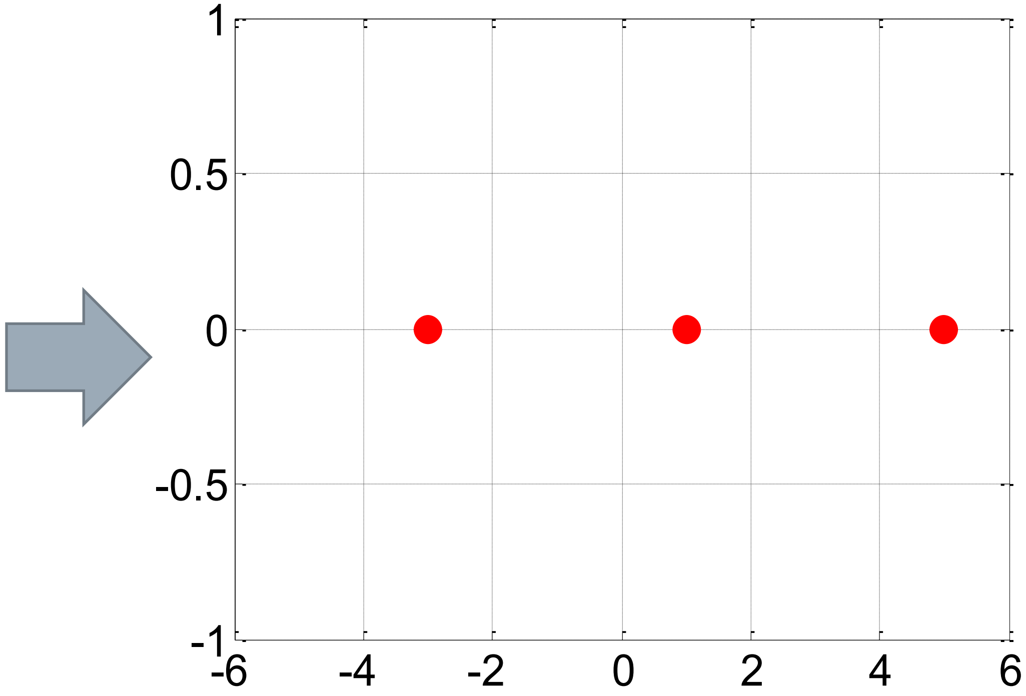
K features
N data points

M features
N data points

Example

$$x = \begin{bmatrix} 0 & 1 \\ 3 & 2 \\ -1 & -2 \end{bmatrix} \qquad w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad Z = \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}$$





Finding the right Transformation

- Let us first forget about reducing the dimensions and focus on reducing the relation between features
- Relation between features is described by the covariance matrix
- $Cov(x, x) = x^T x$
- The best transformation would give no relation between any two features
  - (this means that the covariance matrix an identity matrix)
  - $Cov(z, z) = z^T z = I$

Finding the right Transformation

nxm

Correlated x $\Rightarrow$ z Uncorrelated

M features

nxk

K features

$$Z = xW$$

$$z^T z = I$$

[kxm] [mxn] [nxm] [mxk]

$$z^T z = w^T x^T x w = I$$

Finding the right Transformation

- $cov(x, x)U = U\lambda$
- Where U is the eigen vector and λ is the eigen value
- $U^T cov(x, x)U = \lambda$
- $U^T cov(x, x)U = \lambda^{\frac{1}{2}} I \left(I \quad \lambda^{\frac{1}{2}}\right)^T$
- $\lambda^{\frac{1}{2}^T} \; U^T cov(x, x)U\lambda^{\frac{1}{2}} = I$
- $\left(\lambda^{-\frac{1}{2}}U\right)^T cov(x, x)\left(U\lambda^{-\frac{1}{2}}\right) = I$

Singular Value Decomposition

- $cov(x, x)U = U\lambda$
- $cov(x, x) = U\lambda U^T$
- SVD($cov(x, x)$) $= USV^T$
- Because of the characteristics of covariance matrix
- $U = V^T$

PCA Steps

– Centering each feature around the feature average

$$x_j = x_j - E(x_j)$$

Normalize the data (optional)

$$x_j = \frac{x_j - E(x_j)}{\sigma_j}$$

Calculate the covariance matrix of x

$$cov(x, x) = \frac{1}{n} x x^T$$

Use singular value decomposition to calculate eigen vectors matrix U

If reduction is to be applied pick the first k columns from the u matrix to obtain $U_{reduced}$

Obtain the new transformed data by

$$z = U_{reduced}^T x$$

Note in doing so we ignored the eigen values which represent the significance of each dimension

Obtaining the Original Dimension Approximation

- $z = U_{reduced}^{T} x$
- $x_{approx} = U_{reduced} z$
- In this case the approximation error can be calculated as
- $\frac{1}{m} \sum_{i=1}^{I} \left\| x^{(i)} - x_{approx}^{(i)} \right\|^2$
- Percentage error

$$\frac{\frac{1}{m} \sum_{i=1}^{I} \left\| x^{(i)} - x_{approx}^{(i)} \right\|^2}{\sum_{i=1}^{I} \left\| x^{(i)} \right\|^2}$$

We are interested in a very small percentage of error

Obtaining the Original Dimension Approximation

$$z = U_{reduced}^{T} x \quad \Longleftrightarrow \quad x_{approx} = U_{reduced} z$$

Dr.- Ing. Maggie Mashaly

How to Choose K?

- Iteratively increase k from 1 to n and calculate the error
- Pick the k that ensures an error that is smaller that predefined value
- Luckily the eigen values returned for the case of the covariance matrix provides the contribution of each feature and hence it provides the same value as calculating the error

Thank You

ML