COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

ICAI – GITI/GITT

# PCA
# Estadística II

Eugenio Sánchez Úbeda

January 2024

comillas.edu

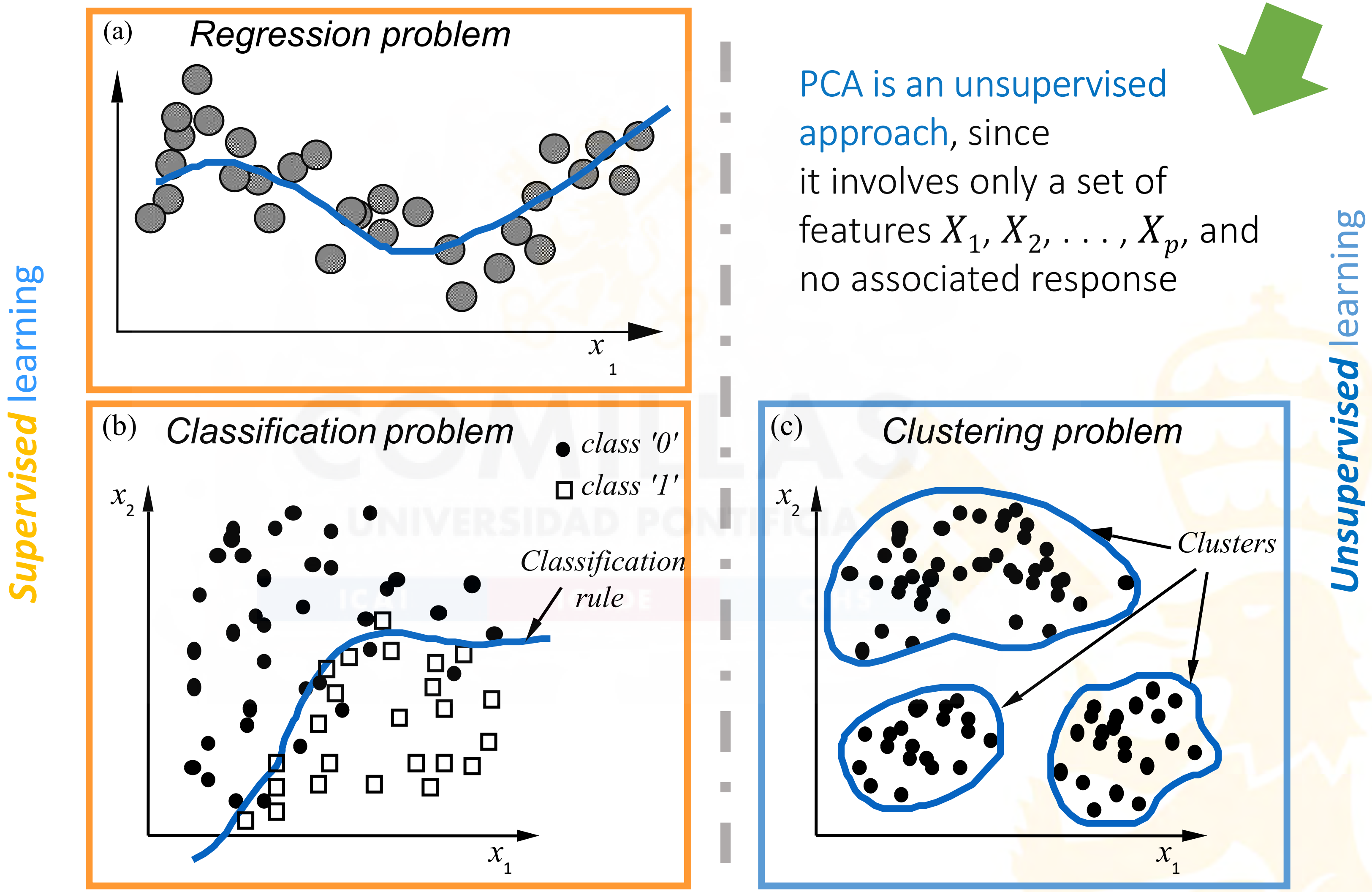
# Bibliography


- G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, *An Introduction to Statistical Learning with Applications in Python*, Springer, 2023 (https://www.statlearning.com)

# 1

# Introduction

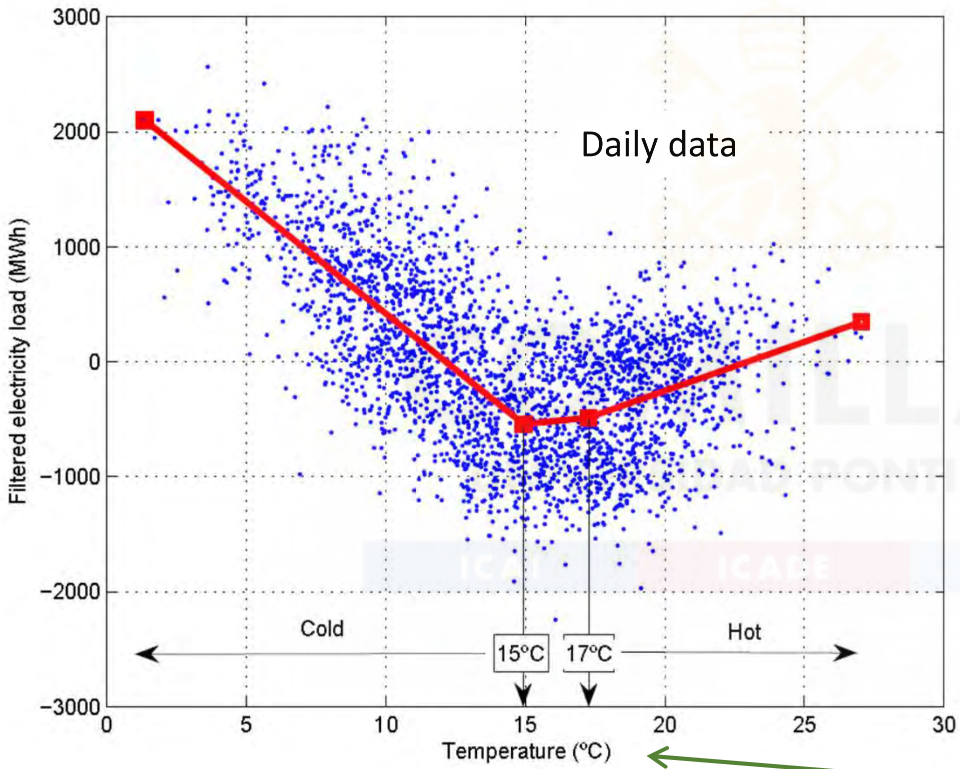
# Principal Components Analysis (PCA)
# Introduction




PCA is an unsupervised approach, since it involves only a set of features $X_1, X_2, \ldots, X_p$, and no associated response

# Principal Components Analysis (PCA)
# Introduction

- Response of electricity demand of a region of Spain to temperature variations




A. Muñoz, E.F. Sánchez-Úbeda, A. Cruz, J. Marín, "*Short-term Forecasting in Power Systems: A guided Tour*", in Handbook of Power Systems II. Pardalos, P.M.; Rebennack, S et al.. Ed. Springer. Berlin, Germany, 2010.

Usually, a regional reference temperature is built as a weighted average of temperatures recorded at different observatories that represent the different climatic regions covering the whole territory of interest

# Principal Components Analysis (PCA) Introduction

- We are interested in a good summary of the temperatures in the form of a small set of new variables


200 weather stations

↓

We would like to find a very low-dimensional representation of the data that captures as much of the information as possible.

↓

Principal Components Analysis (PCA)

# Principal Components Analysis (PCA) Introduction

- PCA aims to determine a few linear combinations of the original variables that can be used to summarize the data set without losing much information
- PCA refers to the process by which principal components are computed and the subsequent use of these components in understanding the data

When faced with a large set of correlated variables, principal components allow us to summarize this set with a smaller number of representative variables that collectively explain most of the variability in the original set.

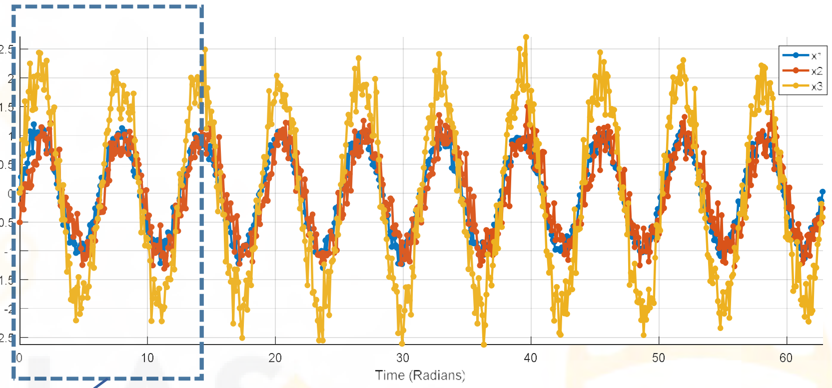
# Principal Components Analysis (PCA)
# What is PCA for?

- PCA is useful for finding out explanatory variables of data that are not directly observed
- This technique is used for Regression, Clustering, and Forecasting when the number of input variables is high and/or variables are correlated
- A way of identifying patterns (driving forces) in data
- After these patterns are found, data can be compressed, reducing the number of dimensions without losing much information (variability)

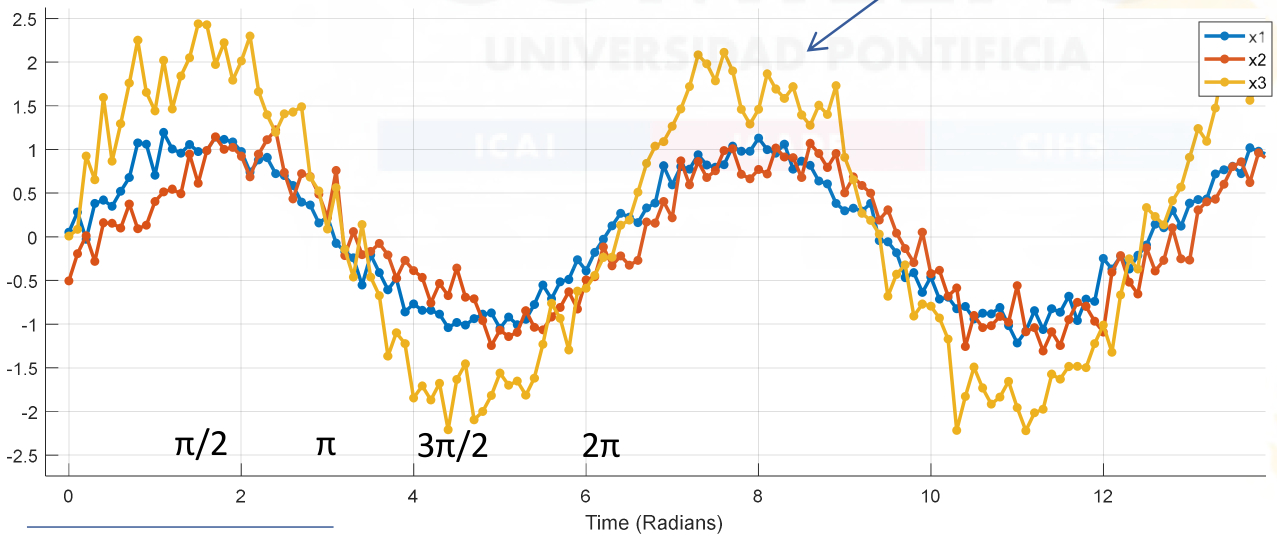
# Principal Components Analysis (PCA) Illustrative example

- We are interested in a good summary of the three variables in the form of a new variable, if possible

```
% CASE 1 WITH DIRECT RELATIONSHIPS

t=[0:0.1:20*pi]';

x1 = sin(t)         + normrnd(0,0.1,size(t));
x2 = sin(t-0.1*pi)  + normrnd(0,0.2,size(t));
x3 = 2 * sin(t)     + normrnd(0,0.3,size(t));
```




Note that they are like daily temperatures (2π is like one year)




- Each variable has a different noise
- $x2$ is a shifted version of $x1$ (same amplitude)
- $x3$ is like $x1$ but with double amplitude

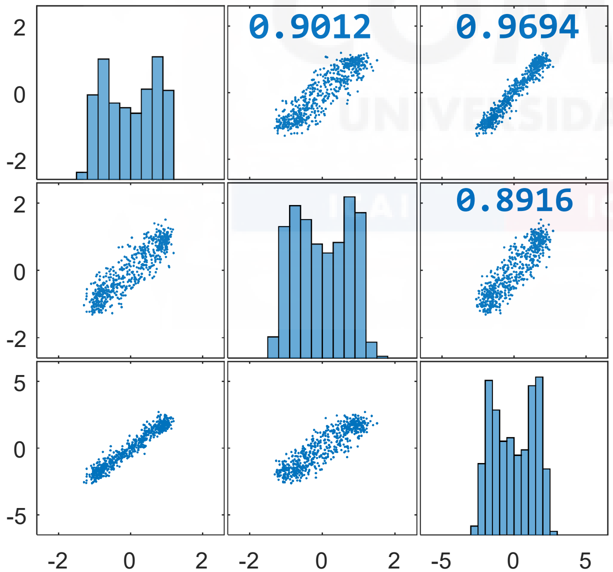
# Principal Components Analysis (PCA)
# Illustrative example

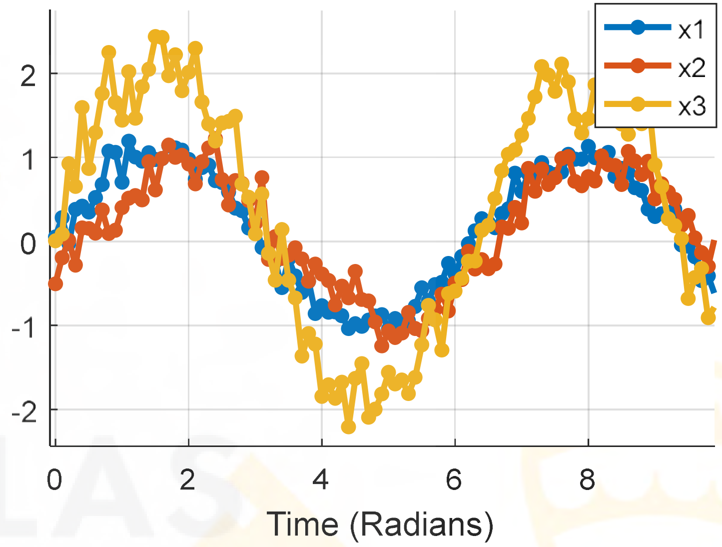
- Data analysis

CASE 1

Nonzero, Positive values

```
cov =
     0.5096    0.4705    0.9974
     0.4705    0.5350    0.9399
     0.9974    0.9399    2.0773

corr =
     1.0000    0.9012    0.9694
     0.9012    1.0000    0.8916
     0.9694    0.8916    1.0000
```



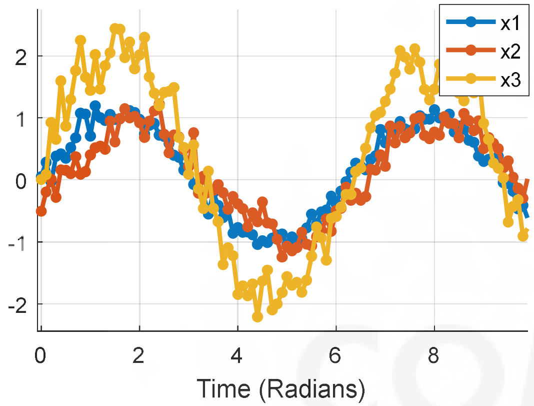



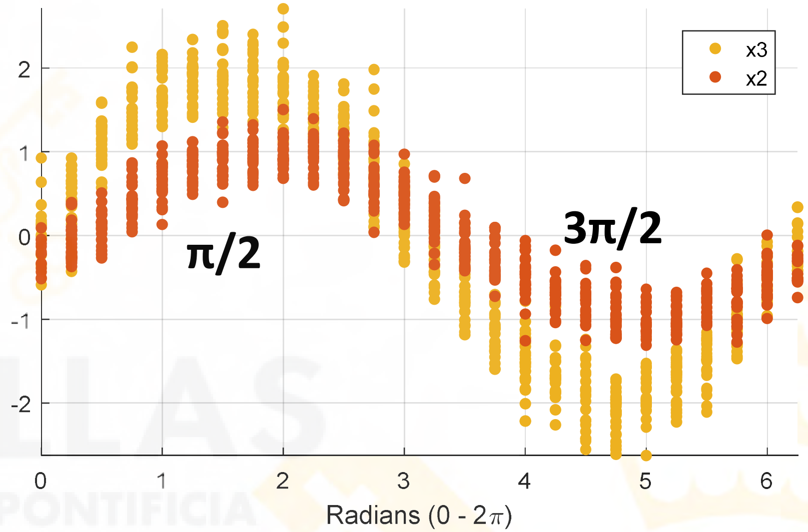

- They are highly correlated
- There are direct relationships between pairs of variables
- According to the histograms, the most frequent values in the extremes of the series seem to be bimodal distributions

# Principal Components Analysis (PCA)
# Illustrative example

- Information about the variables obtained from the sample

CASE 1



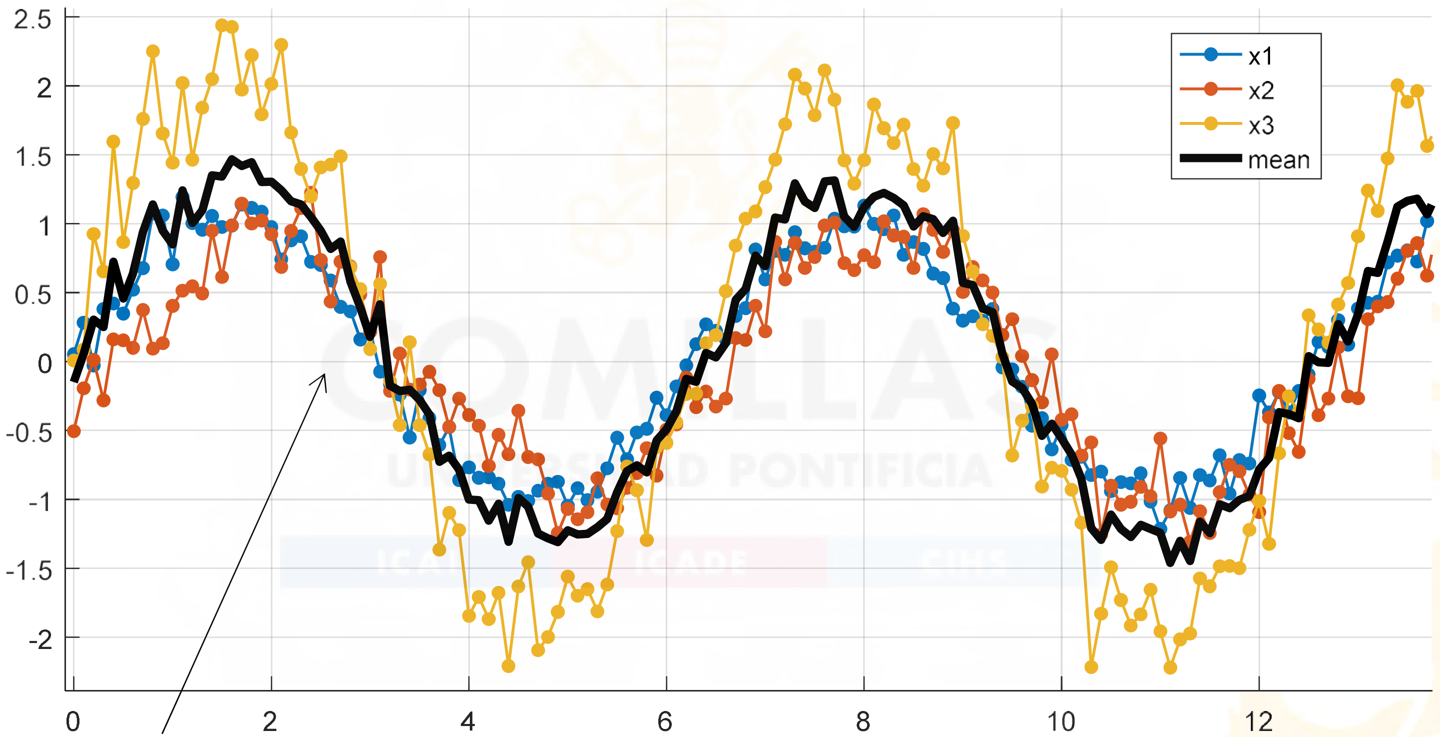

- Their behavior with time is very similar
  - Periodic variables with $t$, with the same period $2\pi$
  - The maximum value is around $\pi/2$ and the minimum is around $3\pi/2$
  - At 0, $\pi$ and $2\pi$ the variables are close to 0

Relevant knowledge: it summarizes qualitatively how these series evolve with time.

# Principal Components Analysis (PCA)
# Illustrative example

- To summarize quantitatively the relevant knowledge, consider the mean of the three variables

CASE 1

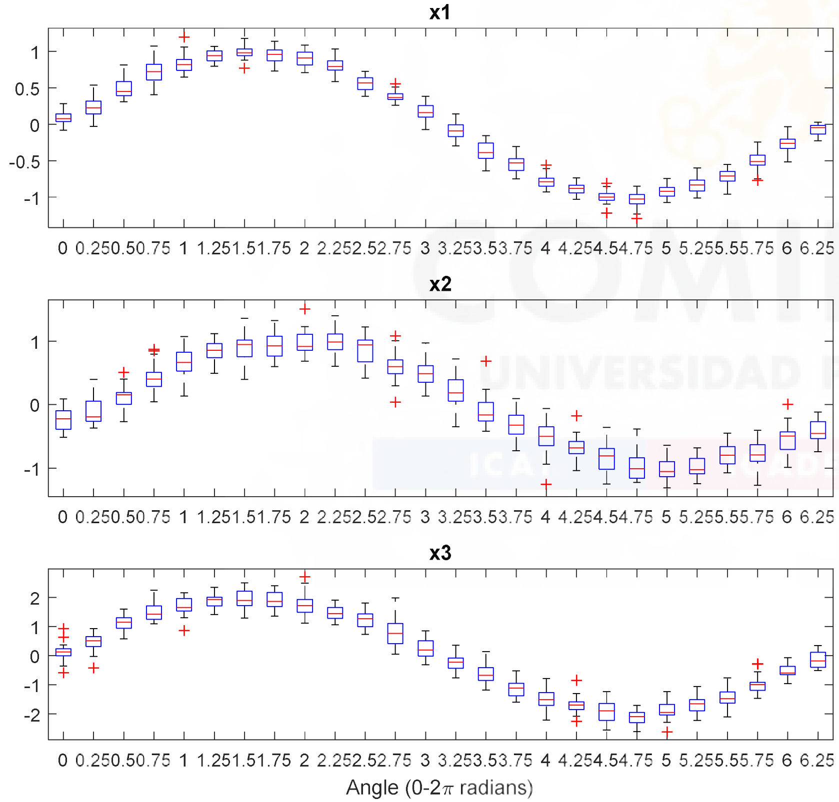



The main dynamics of this new variable are very similar to the original ones.

# Principal Components Analysis (PCA)
# Illustrative example

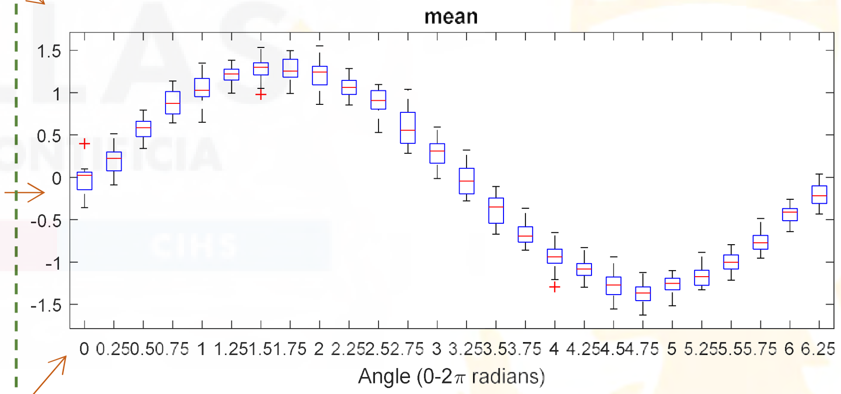
- Information provided by the mean in CASE 1
  - Analyzing the new variable, we can retrieve similar information

**Information about the original variables**




**Information about the new variable**

- Periodic variable with period $2\pi$
- The maximum and minimum values are around $\pi/2$ and $3\pi/2$, respectively
- At 0, $\pi$ and $2\pi$ is close to 0




In this case, the mean can summarize the data set without losing much information.

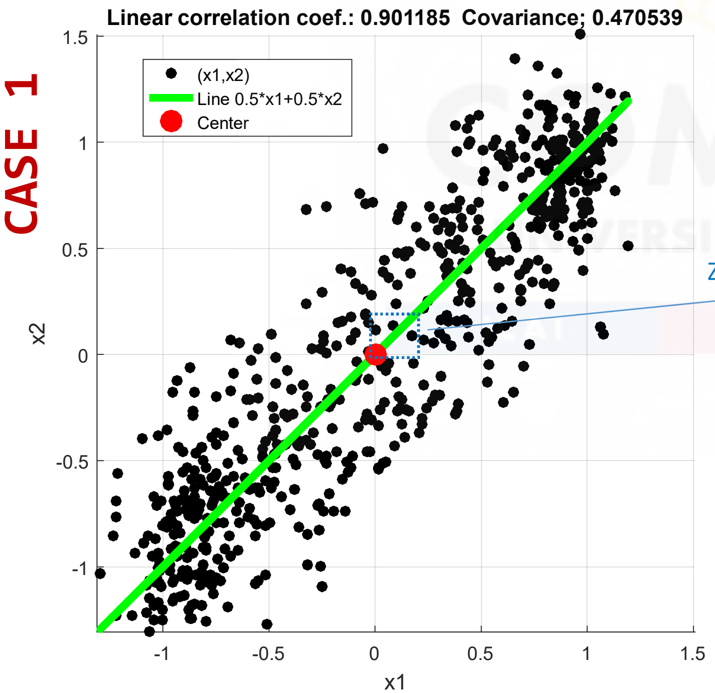
# Principal Components Analysis (PCA)
# Illustrative example

- Considerations about the mean variable
  - The mean is a linear combination of the original variables
  - The mean is obtained by projecting the data onto the direction given by the linear combination in the original input space

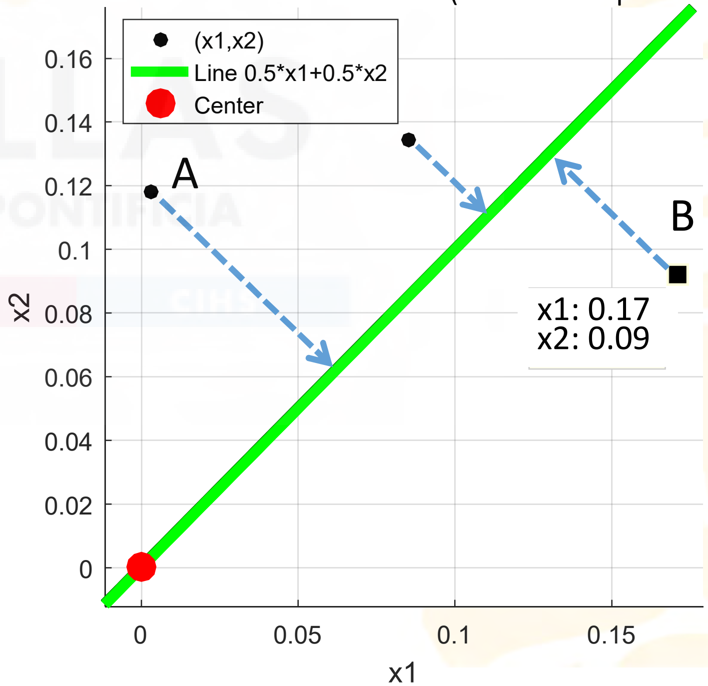
- Example with two input variables:

The mean of point B is larger than the mean of point A (both are positive)

CASE 1




zoom



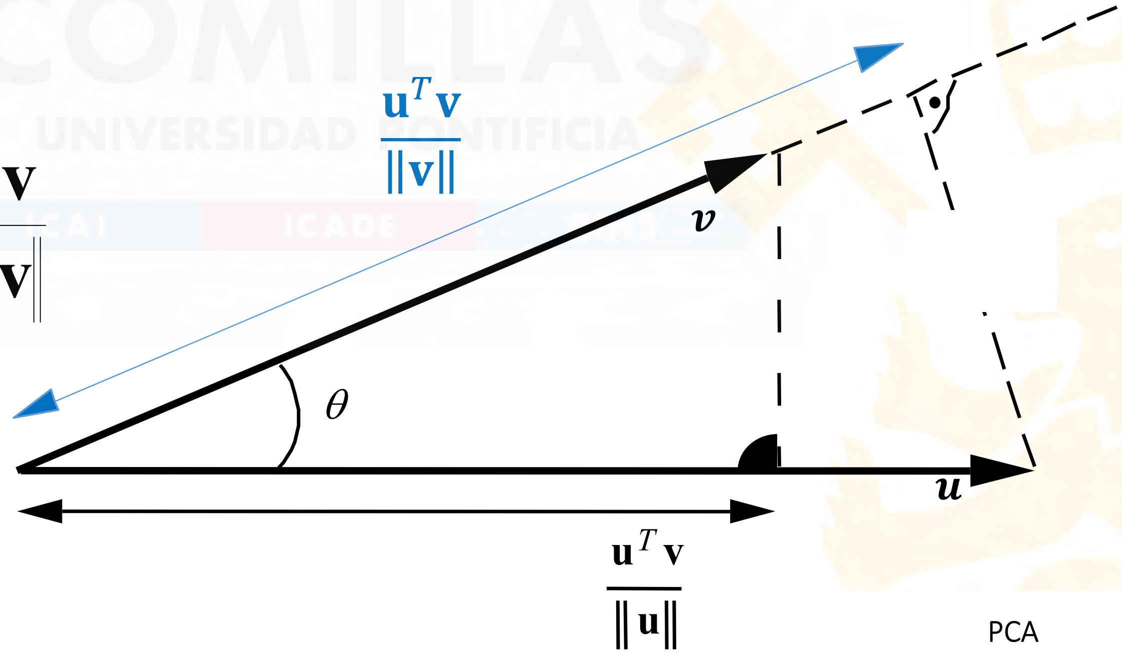
# Principal Components Analysis (PCA)
# Illustrative example

- Considerations about the mean variable
  - Scalar product of two vectors

$$\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = \sum_{k=1,p} u_k v_k$$

- To obtain the standard projection of $\boldsymbol{u}$ along $\boldsymbol{v}$, the previous quantity must be normalized (dividing by the module $\boldsymbol{v}$)

$$\cos\theta = \frac{\mathbf{u}^T\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$$



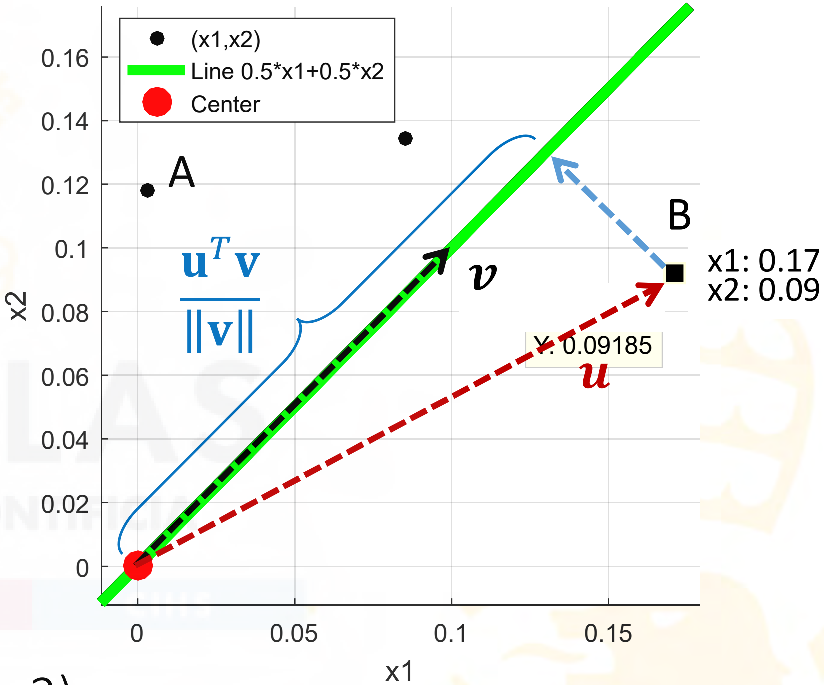
# Principal Components Analysis (PCA)
# Illustrative example

- Projecting a point given by $\boldsymbol{u}$ onto the direction given by $\boldsymbol{v}$

$\mathbf{v} = [\mathbf{0.1}\ \ \mathbf{0.1}]^T \quad \|\mathbf{v}\| = \mathbf{0.14}$

$\mathbf{u} = [\mathbf{0.17}\ \ \mathbf{0.09}]^T$

$$\frac{\mathbf{u}^T\mathbf{v}}{\|\mathbf{v}\|} = \frac{0.1\ 0.17 + 0.1\ 0.09}{0.14} = \mathbf{0.18}$$




$\mathbf{v} = [\mathbf{0.5}\ \ \mathbf{0.5}]^T \quad \|\mathbf{v}\| = \mathbf{0.71}$

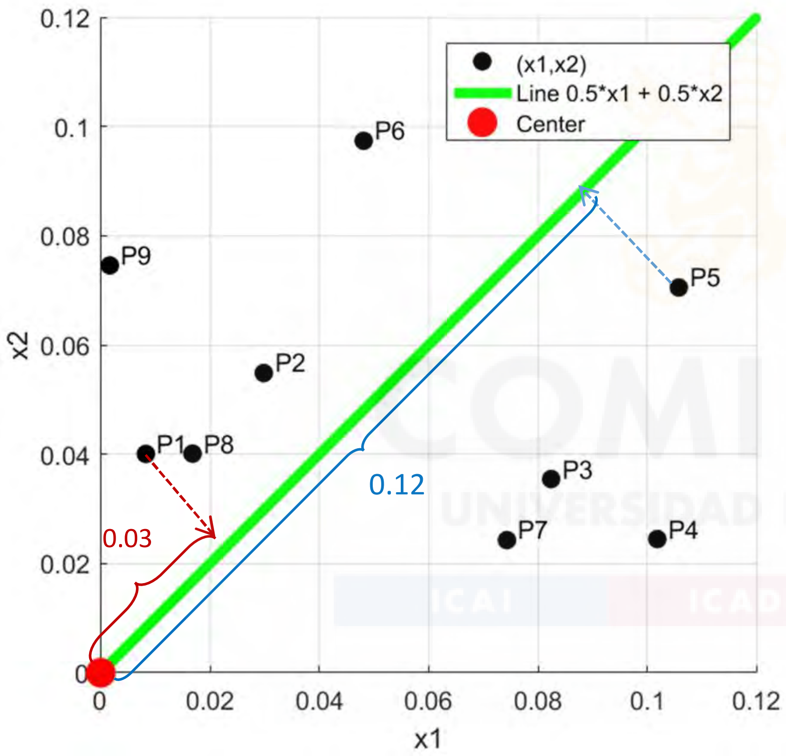
$\mathbf{u} = [\mathbf{0.17}\ \ \mathbf{0.09}]^T$

This is the mean(x1,x2)

$$\frac{\mathbf{u}^T\mathbf{v}}{\|\mathbf{v}\|} = \frac{0.5\ 0.17 + 0.5\ 0.09}{0.71} = \mathbf{0.18}$$

# Principal Components Analysis (PCA)
# Illustrative example

- Projecting a point given by $\boldsymbol{u}$ onto the direction given by $\boldsymbol{v}$



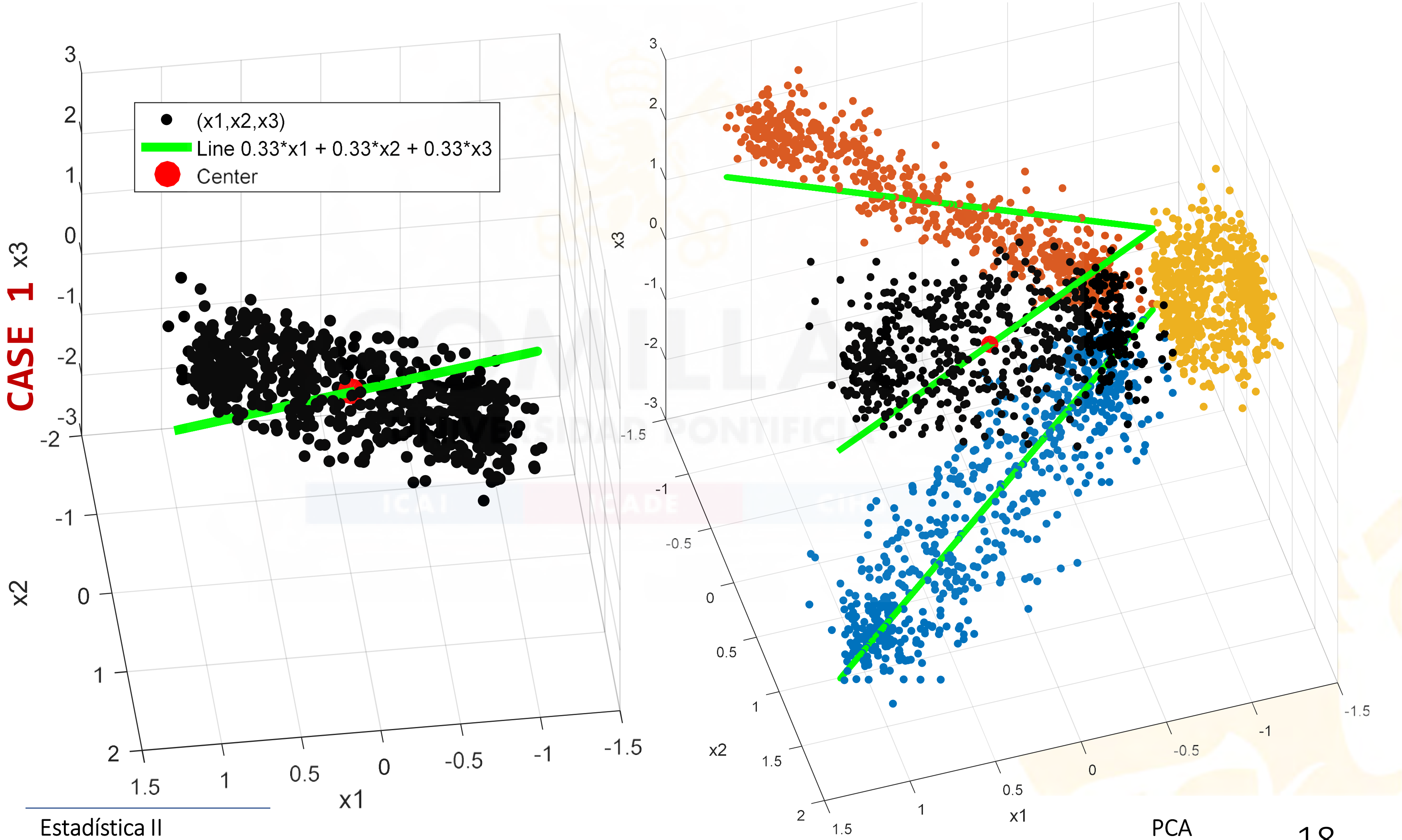

| NAME | x1 | x2 | MEAN | PROJECTION |
|---|---|---|---|---|
| P1 | 0.01 | 0.04 | 0.02 | 0.03 |
| P2 | 0.03 | 0.05 | 0.04 | 0.06 |
| P3 | 0.08 | 0.04 | 0.06 | 0.08 |
| P4 | 0.10 | 0.02 | 0.06 | 0.09 |
| P5 | 0.11 | 0.07 | 0.09 | 0.12 |
| P6 | 0.05 | 0.10 | 0.07 | 0.10 |
| P7 | 0.07 | 0.02 | 0.05 | 0.07 |
| P8 | 0.02 | 0.04 | 0.03 | 0.04 |
| P9 | 0.00 | 0.07 | 0.04 | 0.05 |

Divide the mean by 0.71, i.e., the Euclidean norm of the vector [0.5, 0.5]

Projecting along the **green line** is closely related to computing the mean of the inputs.

# Principal Components Analysis (PCA)
# Illustrative example

- The mean gives the direction with the three inputs

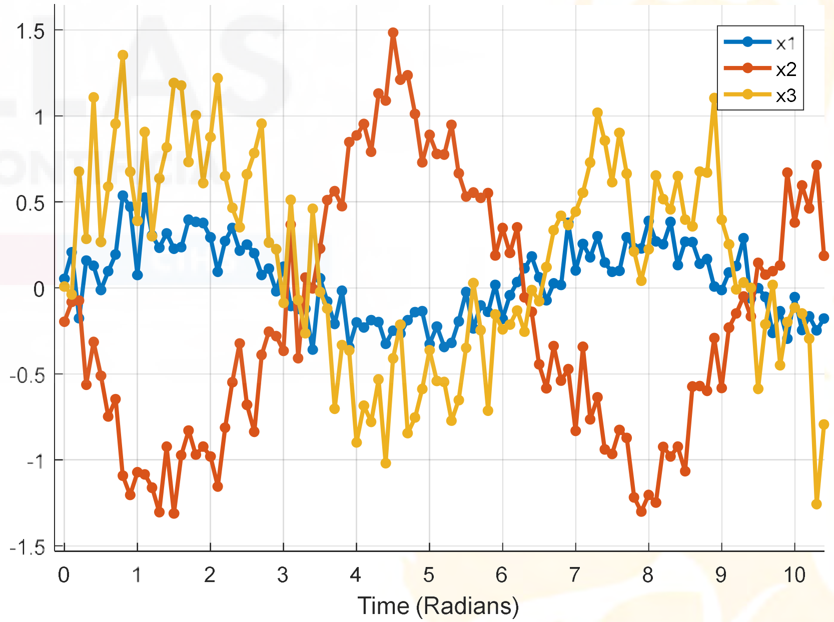


# Principal Components Analysis (PCA)
# Illustrative example

- Further considerations about the mean variable
  - The direction of projection given by the mean is unrelated to the data's joint distribution in the original input space. It has been fixed in advance
  - Maybe the mean is not the main direction that explains most of the variability in the original set, i.e., it is not the best linear combination to reduce the information lost by projecting the observations onto one variable
- Consider the following set of inputs

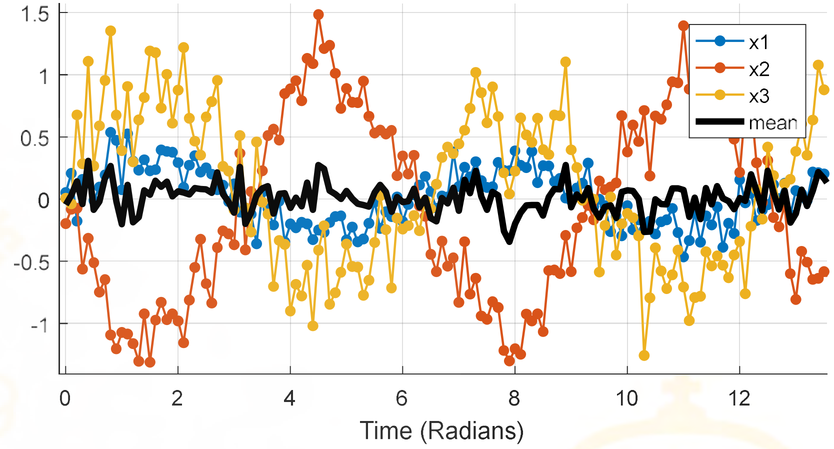
```
% CASE 2 WITH INVERSE RELATIONSHIPS

t=[0:0.1:20*pi]';

x1 = 0.25 * sin(t) + normrnd(0,0.1,size(t));
x2 = sin(t-pi)      + normrnd(0,0.2,size(t));
x3 = 0.75 * sin(t) + normrnd(0,0.3,size(t));
```

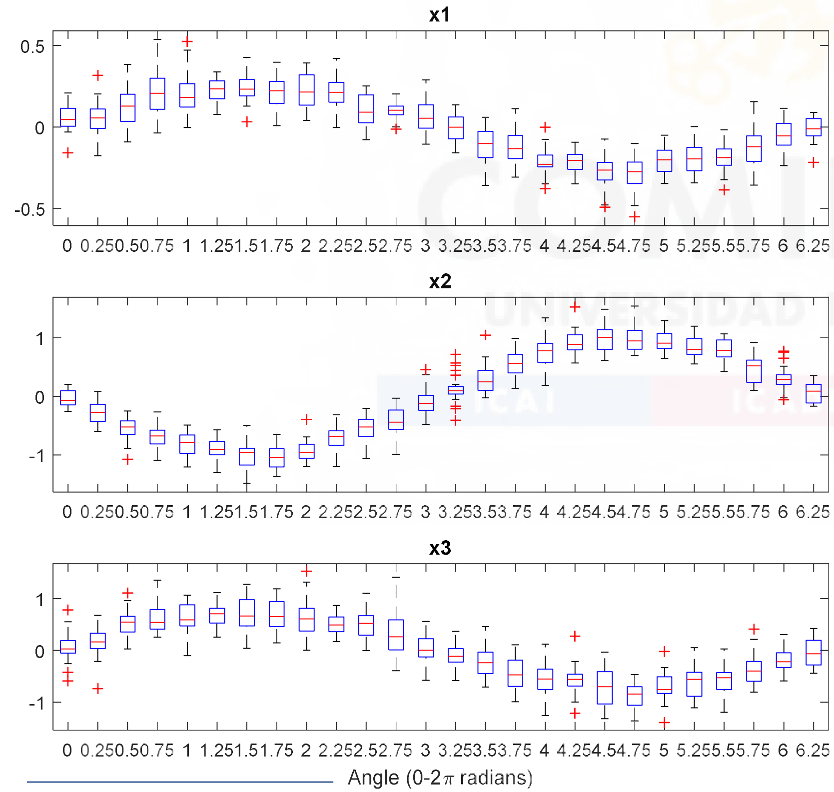
```
cov =
    0.0410   -0.1260    0.0934
   -0.1260    0.5497   -0.3731
    0.0934   -0.3731    0.3657
corr =
    1.0000   -0.8392    0.7627
   -0.8392    1.0000   -0.8323
    0.7627   -0.8323    1.0000
```

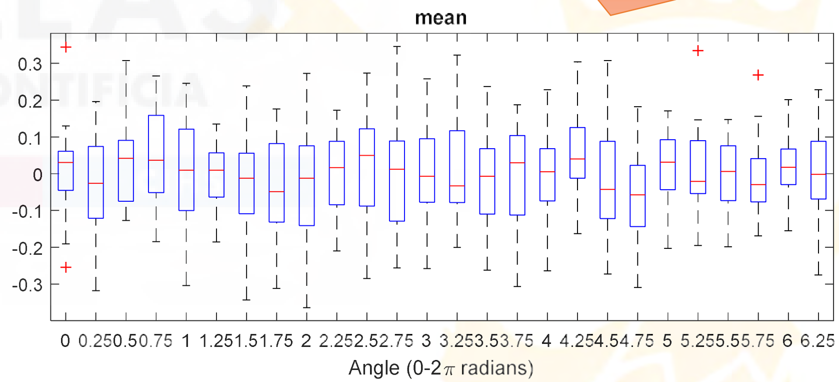


# Principal Components Analysis (PCA)
# Illustrative example

- In CASE 2, the mean variable does not represent the original inputs. Analyzing this variable, we cannot retrieve the main information



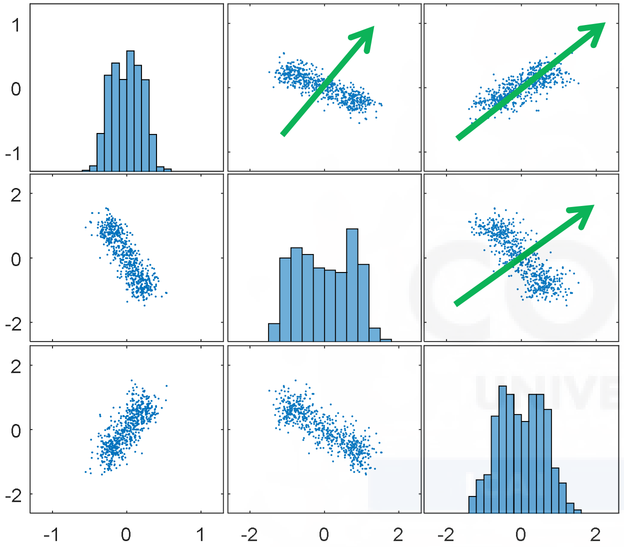



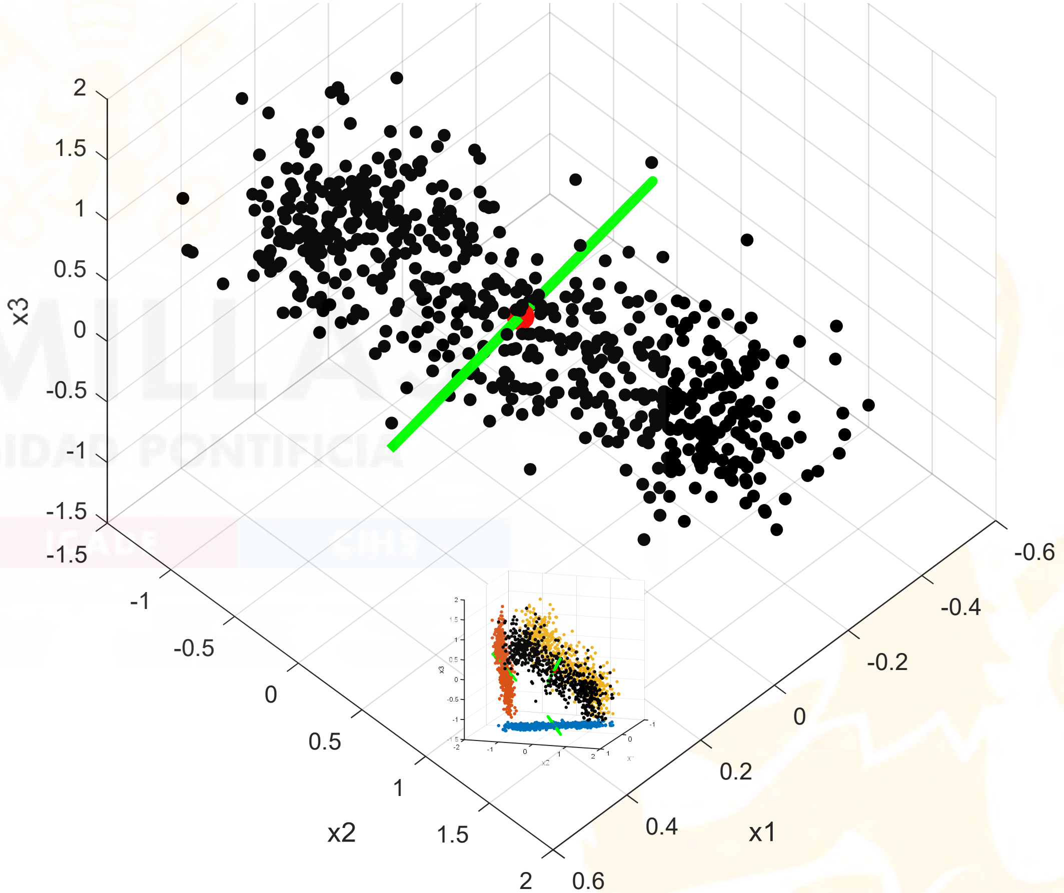




In this case, the mean can't be used to summarize the data set without losing much information.

# Principal Components Analysis (PCA)
# Illustrative example

- In CASE 2, the one-dimensional viewpoint given by the mean is not the most informative of the three-dimensional original space



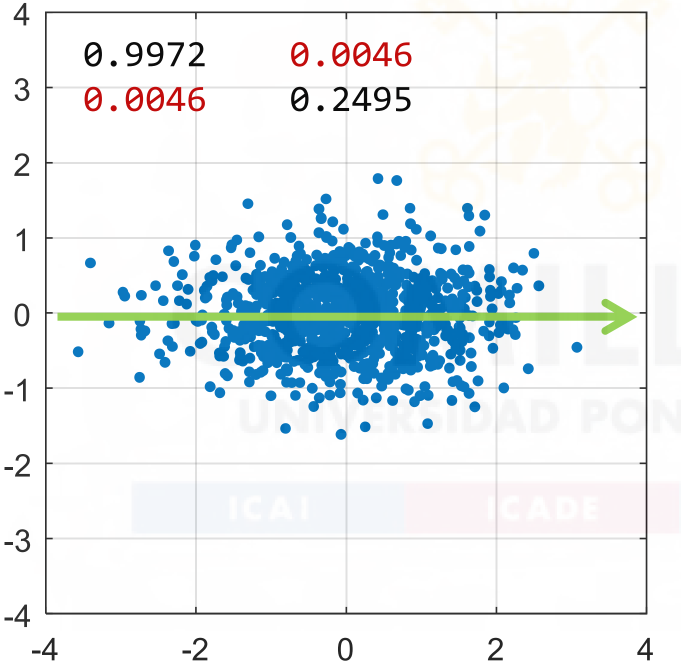




The larger the variance of the projected data, the larger the information you retain in the new variable.

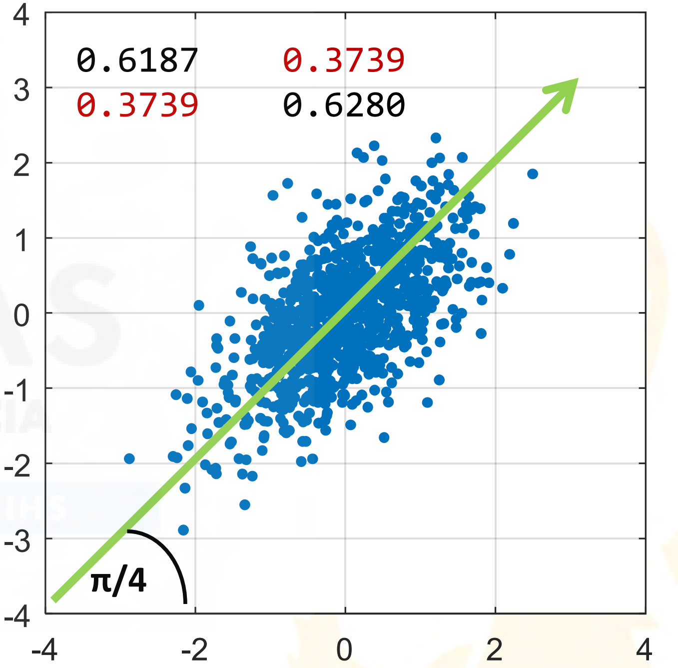
# Principal Components Analysis (PCA)
# The covariance matrix

- The covariance matrix defines the shape of the data. The covariance captures the diagonal spread, while the variance captures the axis-aligned spread

```
n = 1000;
mu = [0 0]; Sigma = [.25 0; 0 1];
Xx = mvnrnd(mu,Sigma,n);
R = [cos(theta) -sin(theta); ...
     sin(theta) cos(theta)];
yy = R*xx';
```




Small covariance, higher variance of $x1$ than $x2$. The spread is mainly along $\boldsymbol{x1}$-axis.

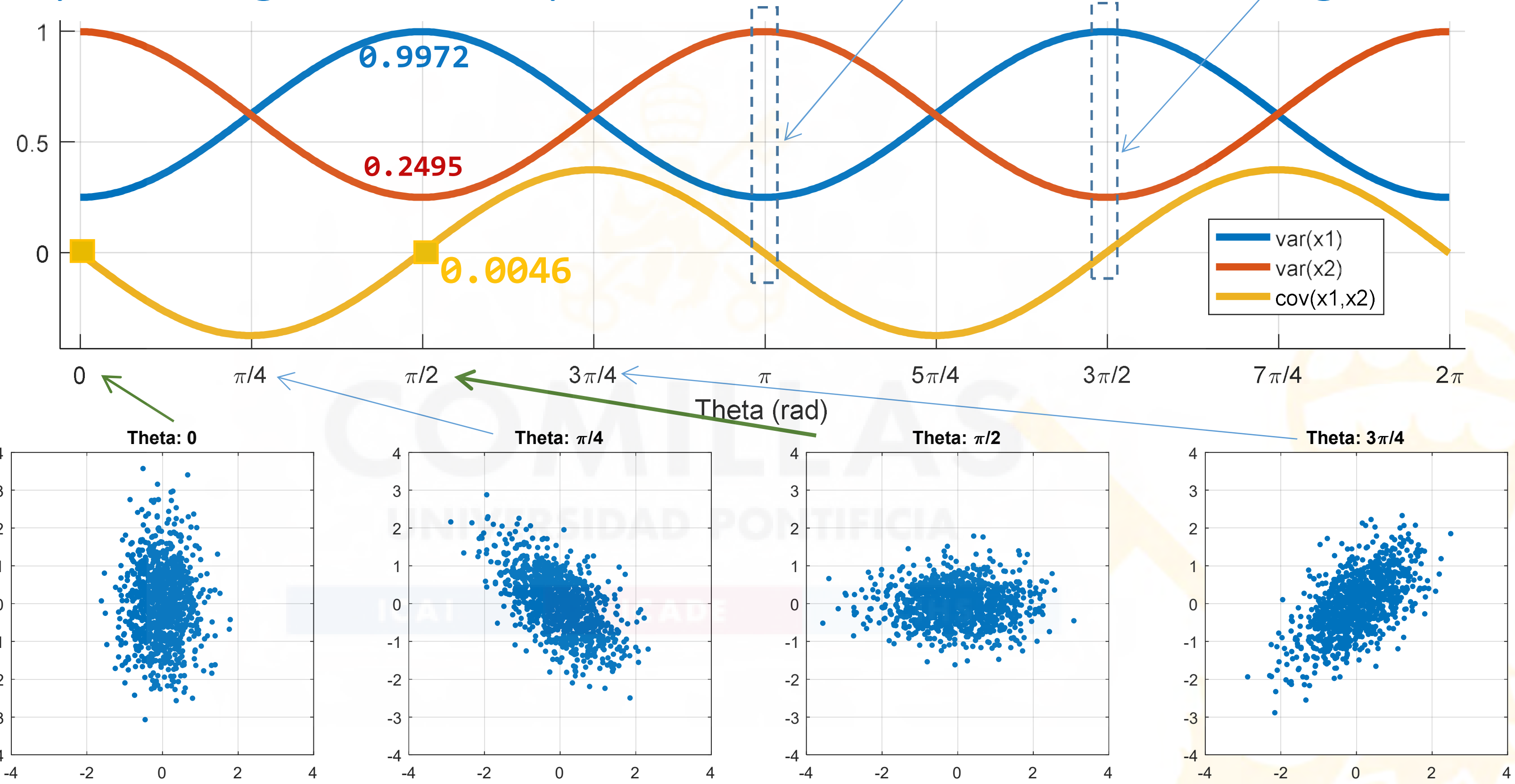



High covariance, variances of $x1$ and $x2$ are very similar. The spread is clearly diagonal.

# Principal Components Analysis (PCA)
# The covariance matrix

When cov($x1$,$x2$) is 0, there are two possible cases: var($x2$) has its maximum value (and var($x1$) the minimum value), or viceversa

- By rotating the scatterplot, the covariance matrix changes

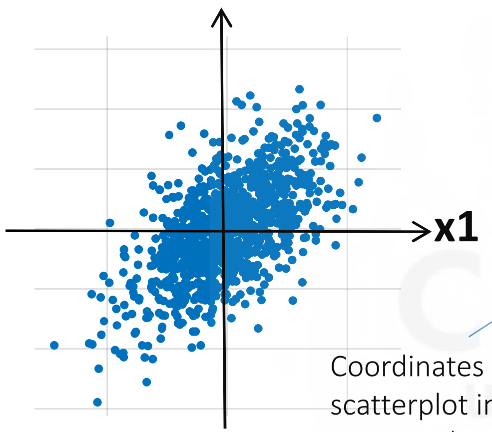



Covariance matrix

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.2495 | -0.0046 | 0.6280 | -0.3739 | 0.9972 | 0.0046 | 0.6187 | 0.3739 |
| -0.0046 | 0.9972 | -0.3739 | 0.6187 | 0.0046 | 0.2495 | 0.3739 | 0.6280 |

# Principal Components Analysis (PCA)
# The covariance matrix

- Instead of rotating the scatterplot, one could rotate the axis and the covariance matrix changes in the same manner



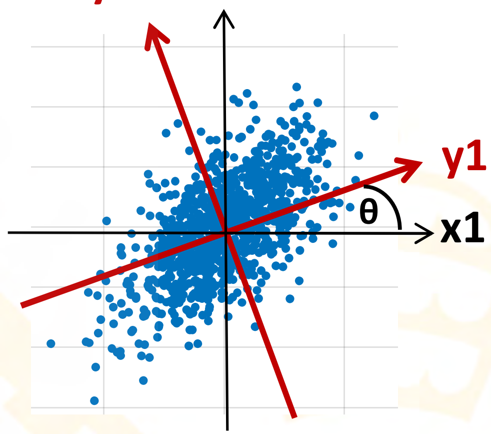

In two dimensions, every rotation matrix has the form:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

$$\mathbf{W} = \begin{bmatrix} 0.9239 & 0.3827 \\ -0.3827 & 0.9239 \end{bmatrix}$$

Coordinates of the scatterplot in the new axis (scores)




$\mathbf{C_{xx}}$ Covariance matrix in the original axis

| | $x1$ | $x2$ |
|---|---|---|
| $x1$ | 0.6187 | 0.3739 |
| $x2$ | 0.3739 | 0.6280 |

$$\mathbf{C_{yy}} = \mathbf{W}\mathbf{C_{xx}}\mathbf{W}^{\mathbf{T}}$$

The rotation matrix allows computing the covariance matrix of the new variables.

$\mathbf{C_{yy}}$ Covariance matrix in the new axis

| | $y1$ | $y2$ |
|---|---|---|
| $y1$ | 0.8844 | 0.2676 |
| $y2$ | 0.2676 | 0.3623 |

# Principal Components Analysis (PCA)
# The covariance matrix

- The larger the variance of the data in the new axis, the larger the information you retain in the new variable
- To maximize the variance in the new axis, we should rotate the reference system such that the covariance matrix $\boldsymbol{C_{yy}}$ is diagonal (i.e., there are no covariance terms in the matrix), the new variables are uncorrelated

$$\mathbf{C_{yy}} = \mathbf{W C_{xx} W^T}$$

Use eigendecomposition to diagonalize the covariance matrix and to obtain the rotation matrix.

- Because the covariance matrix is symmetric and positive semidefinite, its eigenvectors are orthogonal, and all the eigenvalues are all non-negative

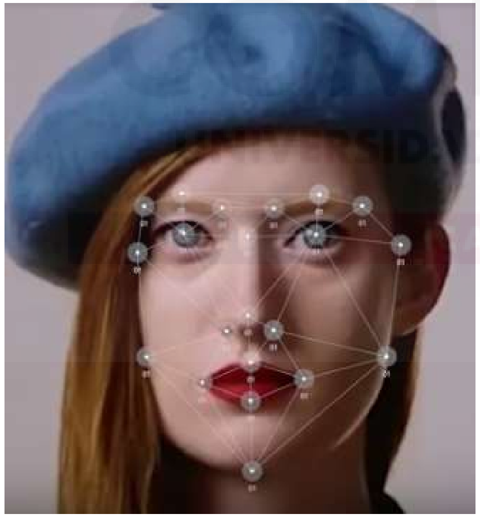
$$\mathbf{C_{xx}} = \mathbf{A \Lambda A^T}$$

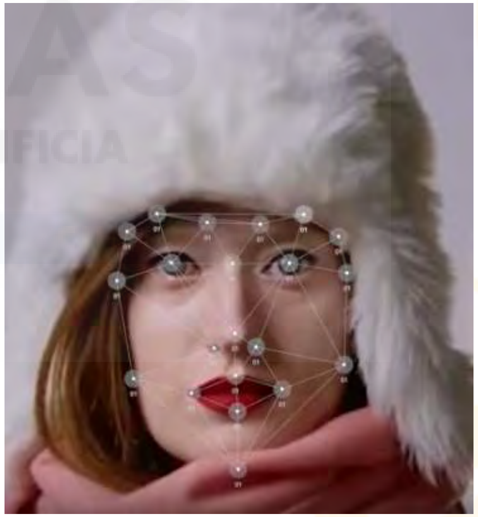
The eigenvectors represent the rotation matrix $A$, and the eigenvalues represent the variance of the data along the eigenvector directions and are in the diagonal matrix $\Lambda$.

# Face recognition

Eigenfaces is the name given to a set of eigenvectors when they are used in the computer vision problem of human face recognition. A set of eigenfaces can be generated by performing a mathematical process called principal component analysis (PCA) on a large set of images depicting different human faces. Informally, eigenfaces can be considered a set of "standardized face ingredients" derived from statistical analysis of many pictures of faces. Any human face can be considered to be a combination of these standard faces.

Source: Wikipedia





Photos: J. Rodríguez Reconocimiento facial: cuando tu cara es tu nuevo DNI El País Dic 2017

Biometric identifiers, such as the distance between eyes, skin color, and the length of the person's nose.

2

# PCA Approach

# PCA approach
# Overview

- Given a set of original variables (usually correlated)

$$X_1, X_2, X_3, X_4, \cdots, X_p$$

- The aim of PCA is to determine the principal components, a new set of uncorrelated variables ($M \leq p$)

$$Z_1, Z_2, \cdots, Z_M$$

- Where each principal component is a linear combination of the original variables

Graphically, PCA rotates the original axis to the directions of maximum variability.

$$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j$$

$$\sum_{j=1}^{p} \phi_{j1}^2 = 1$$

First principal component:

$$Z_1 = \phi_{11} X_1 + \phi_{21} X_2 + \ldots + \phi_{p1} X_p$$

# PCA approach Overview

These directions are named as Principal Directions of the data and the new values using Principal Directions to define data are named Principal Components

- Terminology

Principal components → $$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j$$ ← Original variables

Loadings (the weight by which each standardized original variable should be multiplied to get the component score)

$$\phi_1 = (\phi_{11} \ \phi_{21} \ \ldots \ \phi_{p1})^T$$

Component scores (the transformed variable values corresponding to a data point)

$$z_{i1} = \phi_{11} x_{i1} + \phi_{21} x_{i2} + \ldots + \phi_{p1} x_{ip}$$

# PCA approach
# Overview

- Estimation of the loadings (i.e., the coeff. of the model)

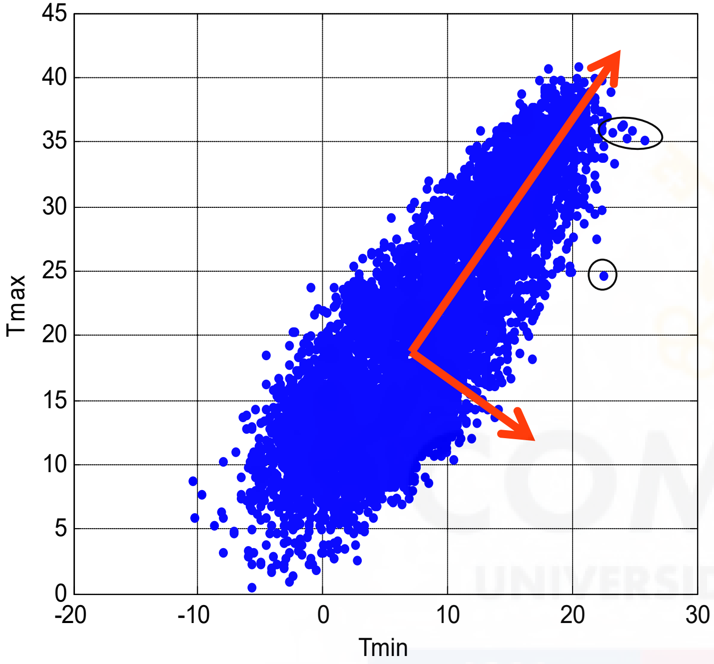
$$Z_m = \sum_{j=1}^{p} \phi_{jm} X_j$$

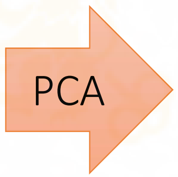
In general matrix $\boldsymbol{X}$ (and consequently the covariance matrix) has rank $\boldsymbol{p}$, then there are as many Principal Components as original variables

Find the orthogonal transformation (the loadings) to obtain a new set of linearly uncorrelated variables such that the first principal component has the largest possible variance and each succeeding component, in turn, has the highest variance possible under the constraint that it is orthogonal to the preceding components

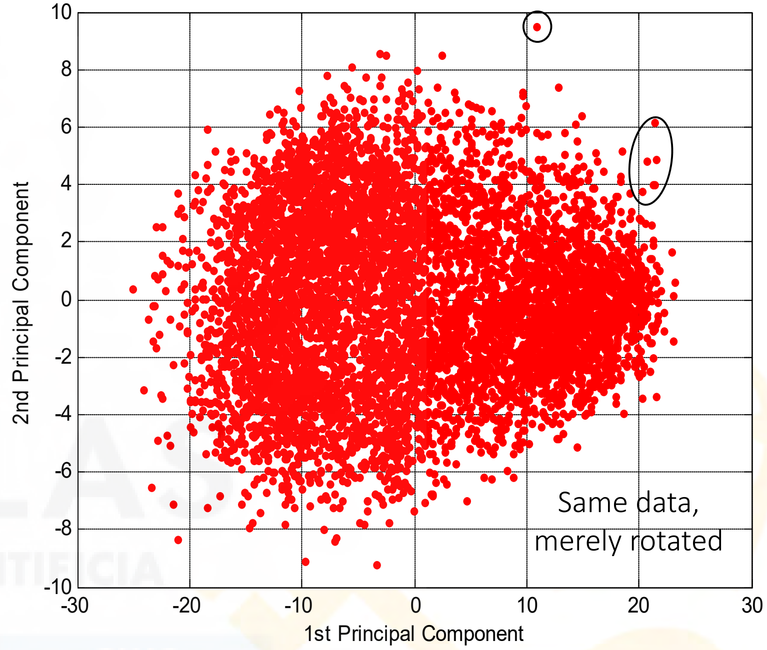
- Algorithm
  - For each variable, its mean is subtracted, and then the covariance matrix is computed
  - Compute eigenvectors and eigenvalues of the covariance matrix
  - Sort the eigenvectors with respect to the eigenvalues (higher to lower)
  - *Apply the linear transformation to the data to obtain the scores*

# PCA approach
# Example with Tmin and Tmax



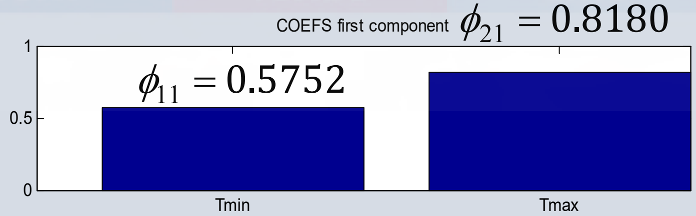




Pca1 = 0.5752 Tmin + 0.8180 Tmax




## Eigenvectors (loadings)

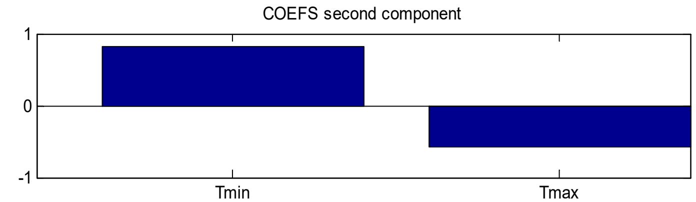
0.5752 0.8180




## Eigenvalues (variances)

116.2682

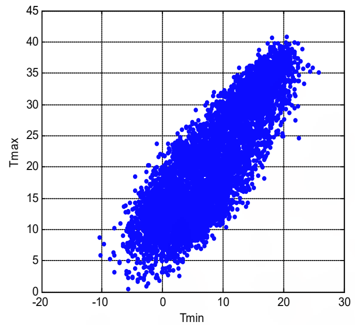
0.8180 -0.5752



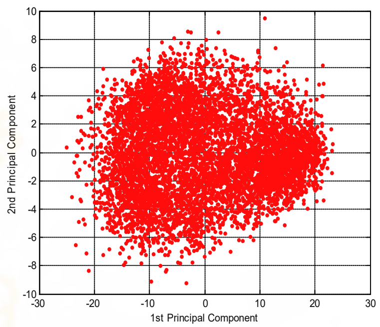

8.8544

# PCA approach
# Example with Tmin and Tmax

- Variances before and after PCA




PCA




The sum of variances of the $\boldsymbol{p}$ Principal Components is equal to the sum of variances of the $\boldsymbol{p}$ original variables and equal to the sum of eigenvalues.

- Covariance matrices

| | Tmin | Tmax |
|---|---|---|
| Tmin | 44.3886 | 50.5390 |
| Tmax | 50.5390 | 80.7340 |

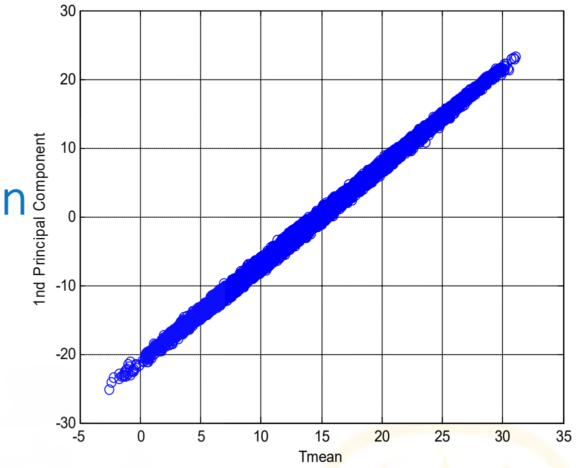
| | pc1 | pc2 |
|---|---|---|
| pc1 | 116.2682 | 0.0000 |
| pc2 | 0.0000 | 8.8544 |

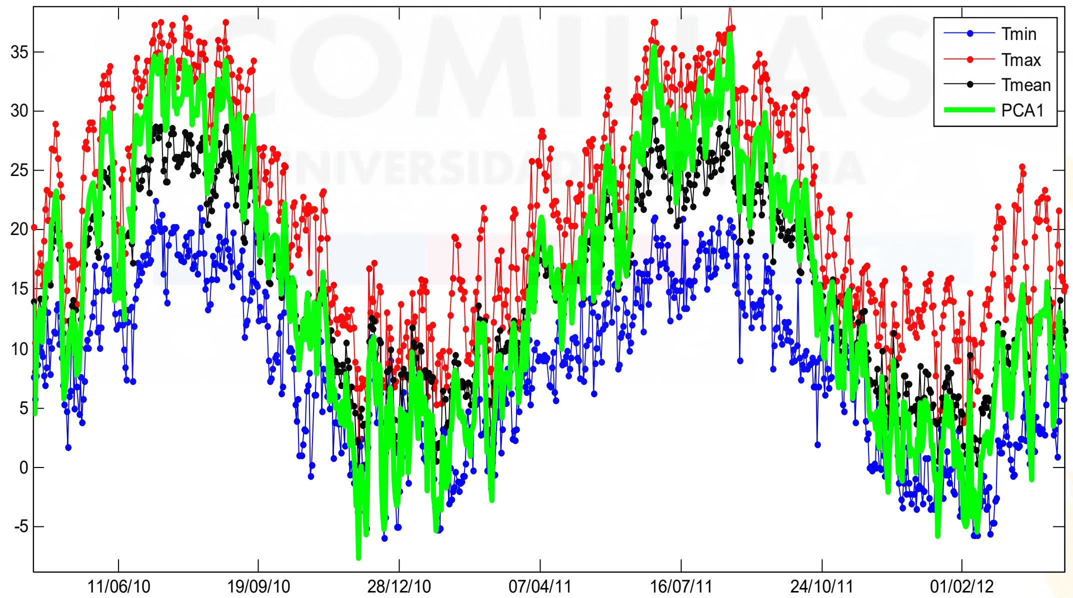
- There is variance in each variable Tmin and Tmax
- There is also covariance
  - cov(Tmin,Tmax) = 50.539
  - Tmin and Tmax are correlated

- Principal components are uncorrelated
  - cov(pc1,pc2) = 0
  - They “remove” the covariance in the new basis

# PCA approach
# Example with Tmin and Tmax

- Pc1 vs. Tmean
  - Highly correlated
  - But it is not the same, pc1 has more information
  - The variability of pc1 is higher (direction of maximum variability)
- Tmean vs. pc1 + mean(Tmean)



# PCA
# The Proportion of Variance Explained

- Although there are $p$ principal components, we usually are not interested in all of them
- We would like to use the smallest number of principal components required to get a good understanding of the data

⬇

- How much of the information in each data set is lost by projecting the observations onto the first few principal components?

⬇

Compute the proportion of variance explained by each principal component and use the scree plot.

# PCA
# The Proportion of Variance Explained

- The total variance present in a data set of $n$ observations

$$\sum_{j=1}^{p} \text{Var}(X_j) = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} x_{ij}^2$$

Assuming that the variables have been centered to have a mean zero.

- The variance explained by the $m$-th principal component is

$$\frac{1}{n} \sum_{i=1}^{n} z_{im}^2 = \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{p} \phi_{jm} x_{ij} \right)^2$$

It is the eigenvalue of this principal component.

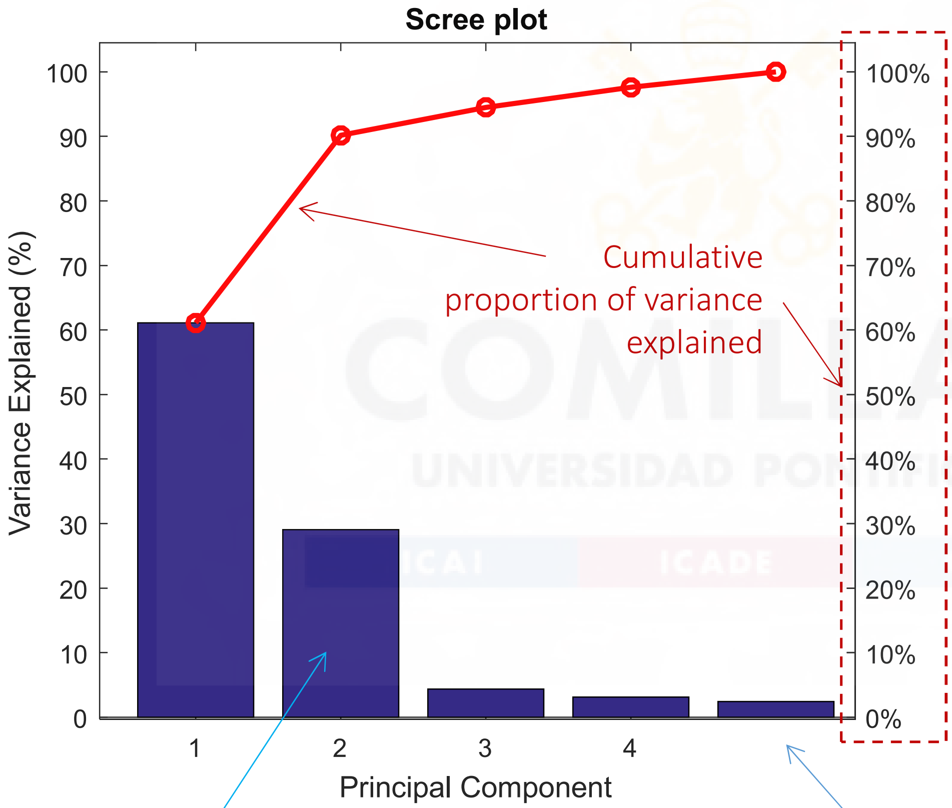
- And the proportion of variance explained (PVE) by the $m$-th component

$$\frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{p} \phi_{jm} x_{ij} \right)^2}{\sum_{j=1}^{p} \sum_{i=1}^{n} x_{ij}^2}$$

- The PVEs are positive
- Their sum is one

# PCA
# How Many Principal Components to Use

- We typically decide by examining the scree plot




PVE (%) of the principal components

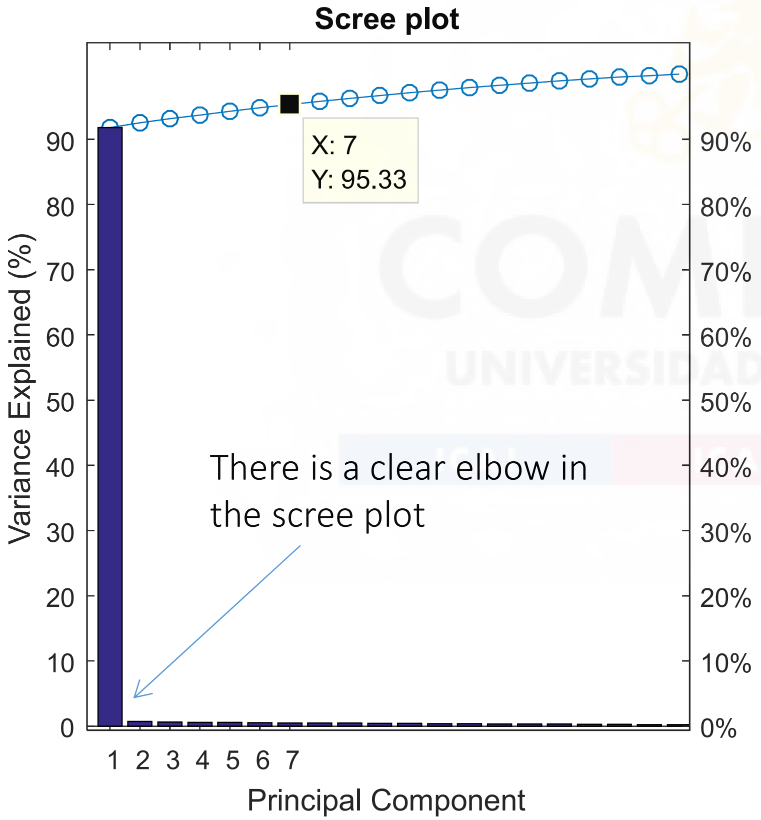
Using all components, the variance is entirely explained (100%)

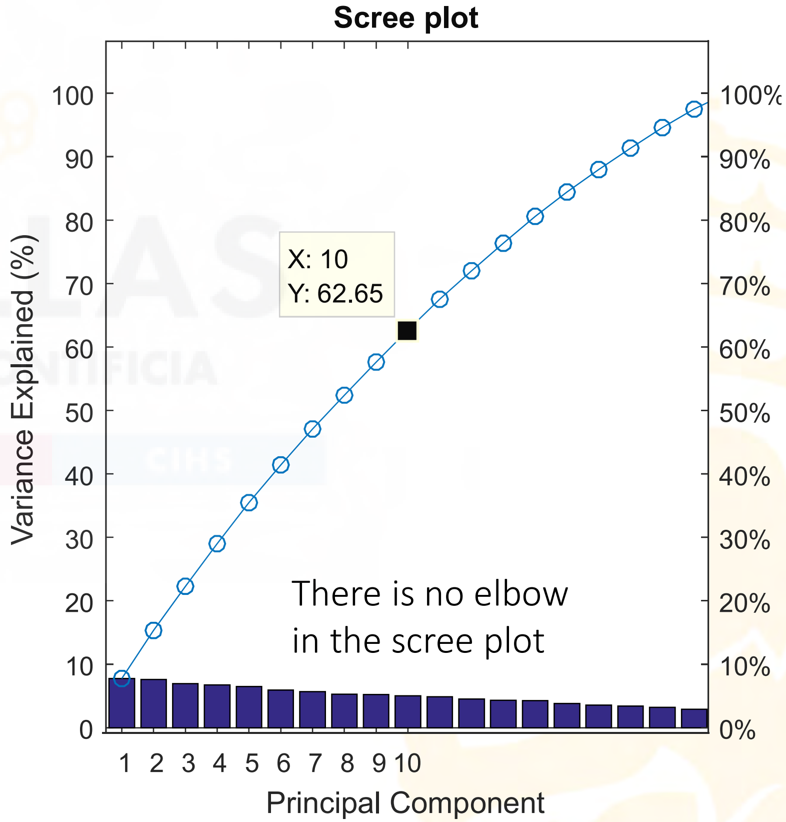
- We choose the smallest number of principal components that are required to explain a sizable amount of the variation in the data
- This is done by eyeballing the scree plot and looking for an elbow, i.e., a point at which the proportion of variance explained by each subsequent principal component drops off

# PCA
# How Many Principal Components to Use

- Extreme cases examining the scree plot

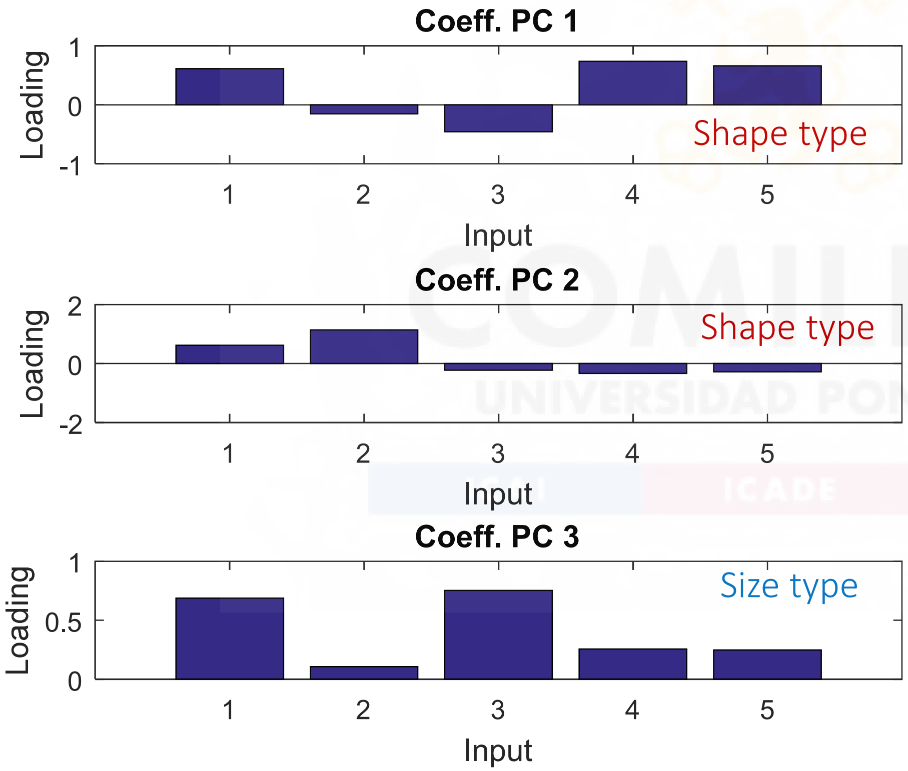
| Highly correlated inputs $n$ = 200; $p$ = 20; corr(all)=0.9 | Uncorrelated inputs $n$ = 200; $p$ = 20; |
|---|---|



# PCA
# Interpretation of principal components

- The loadings summarize the relationship between the original inputs and the principal components




- When the eigenvector of the principal component has positive and negative coordinates, the type of component is a shape type. Coordinates can be grouped into two sets of different signs. The weighted coordinates of a variable set are opposed to the weighted coordinates of the other set
- When there is a high positive correlation among all variables, the eigenvector of the principal component has all coordinates with the same sign. This component is a size type. This component can be interpreted as a weighted average of all original variables.

# PCA
# Scaling the variables

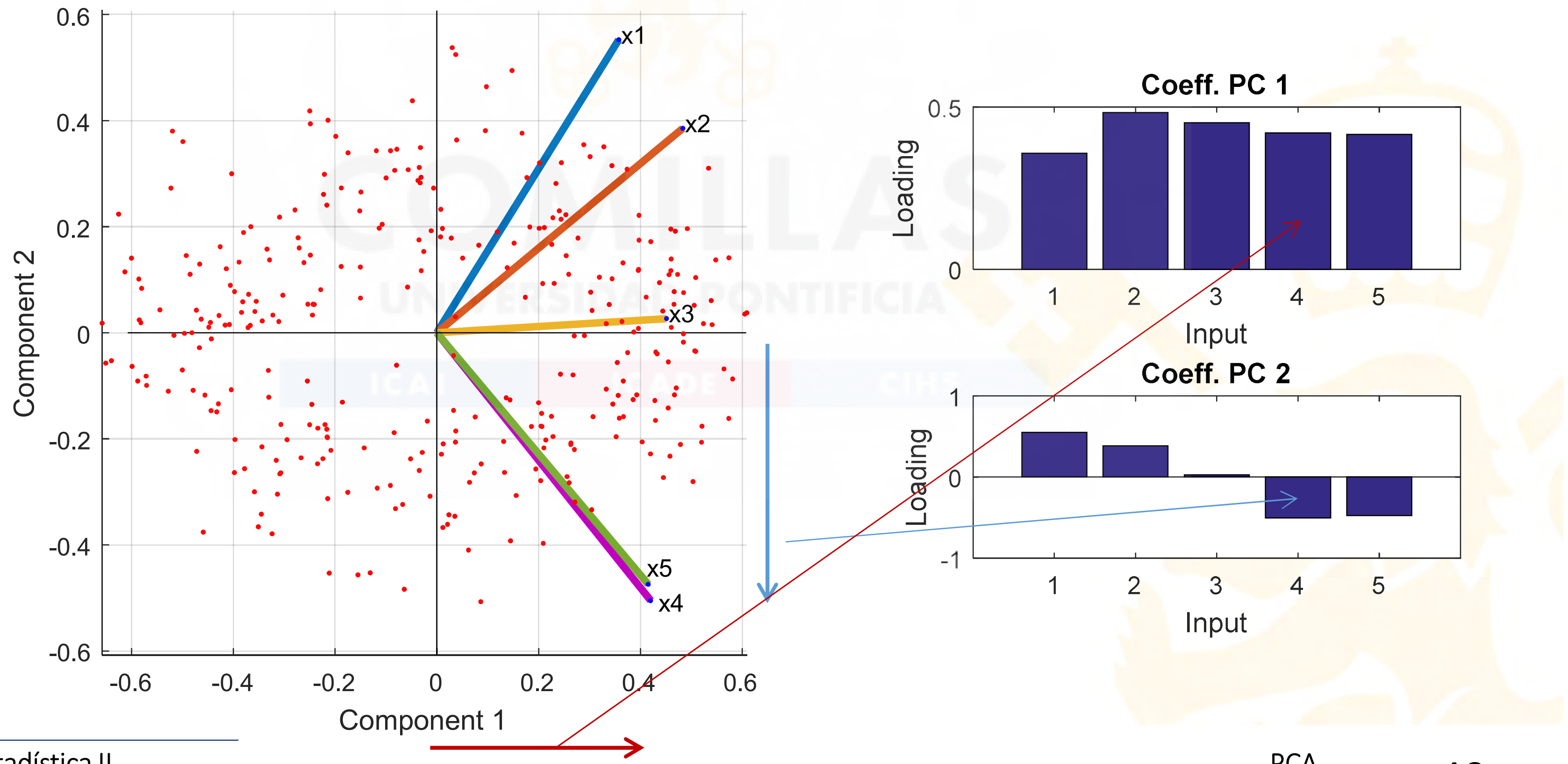
- PCA is sensitive to the relative scaling of the original variables
  - The results of the analysis will depend on what units of measurement are used to measure each variable
- It gives more importance to those variables that have higher variances than to those variables that have very low variances

↓

- If the mean is very different on each variable, decide to scale each variable to analyze all of them in similar ranges
- If the variability is very different, decide whether to standardize the variation or not. Performing this standardization hides the impact of the inherent variation of each variable.
- Standardizing the data before performing the principal component analysis gives more importance to the correlations among variables

# PCA
# Biplot graph

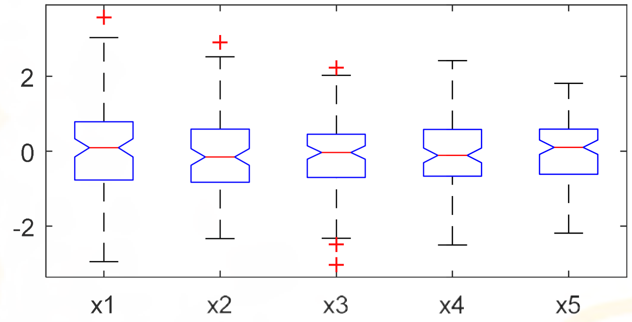
- The Biplot graph allows visualizing both the orthonormal principal component coefficients for each variable and the principal component scores for each observation in a single plot



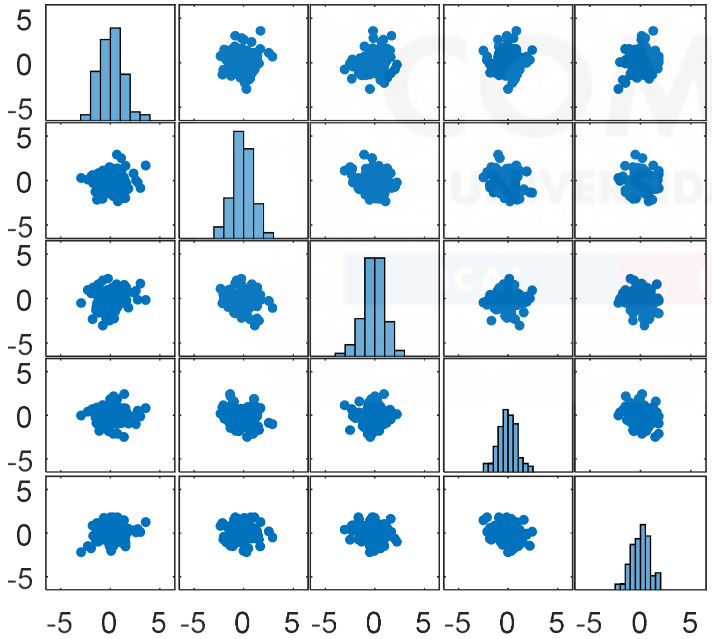
# PCA
# Illustrative synthetic cases

- C1: Uncorrelated and standardized input variables

```
n = 100; p = 5;
X = normrnd(0,1,n,p);
```



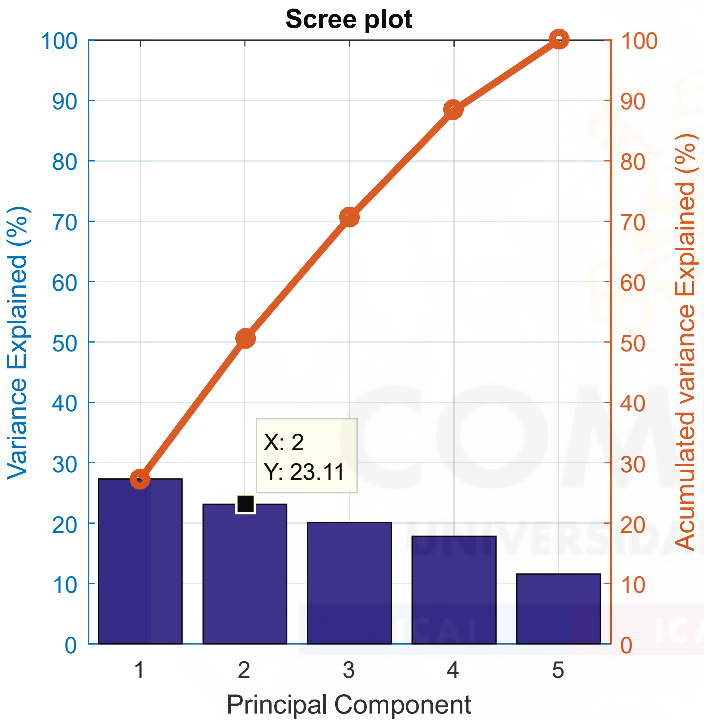

Rows of $X$ correspond to observations, and columns correspond to variables.

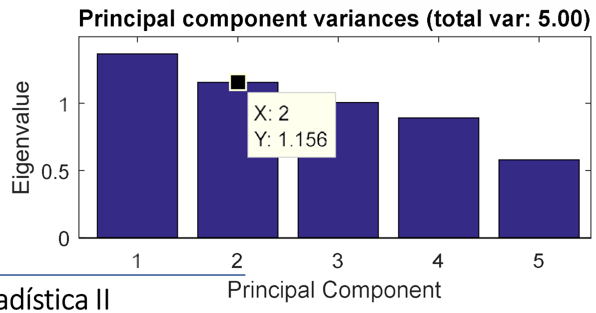



```
cov(X)=
    1.3512    0.0881    0.1556    0.0810    0.1621
    0.0881    1.0101   -0.1868   -0.1287    0.0218
    0.1556   -0.1868    1.0329    0.0382   -0.1045
    0.0810   -0.1287    0.0382    0.8873   -0.1868
    0.1621    0.0218   -0.1045   -0.1868    0.7632

corr(X)=
    1.0000    0.0754    0.1317    0.0740    0.1596
    0.0754    1.0000   -0.1829   -0.1360    0.0248
    0.1317   -0.1829    1.0000    0.0400   -0.1177
    0.0740   -0.1360    0.0400    1.0000   -0.2270
    0.1596    0.0248   -0.1177   -0.2270    1.0000
```

# PCA
# Illustrative synthetic cases

- ## C1: Uncorrelated and standardized input variables




PVE(2) % = 100 * 1.156/5.0 = 23.11 %

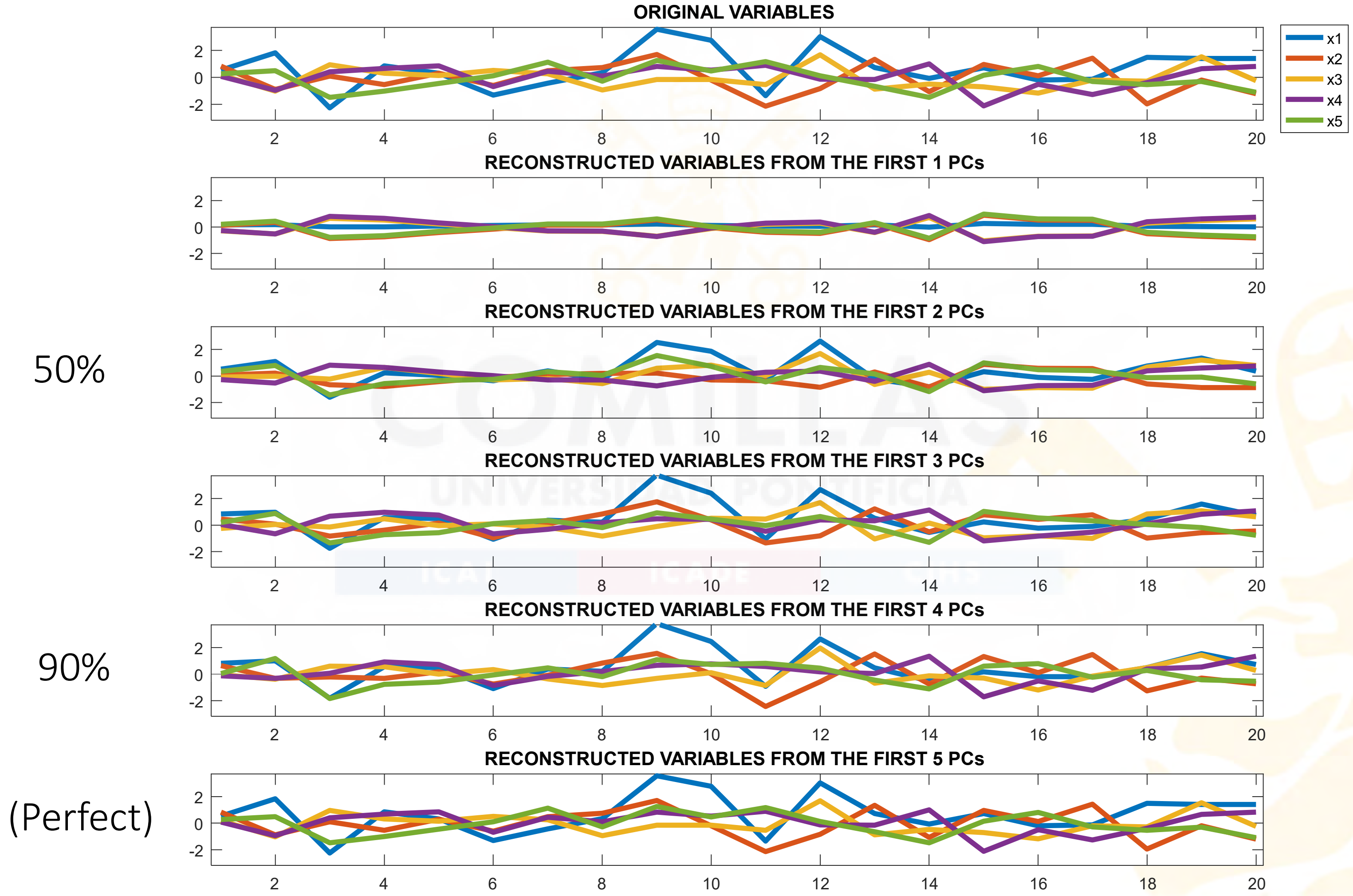
```
cov(Z)=
     1.3669   -0.0000    0.0000    0.0000   -0.0000
    -0.0000    1.1556   -0.0000   -0.0000   -0.0000
     0.0000   -0.0000    1.0064   -0.0000    0.0000
     0.0000   -0.0000   -0.0000    0.8908    0.0000
    -0.0000   -0.0000    0.0000    0.0000    0.5803
```

- The first few principal variables do not explain most of the variability in the original set $X$
- Four principal components are needed to explain around 90% of the variability
- The total variance is spread out linearly along the 5 principal variables

In this problem, PCA does not allow reducing the number of dimensions without losing much information.

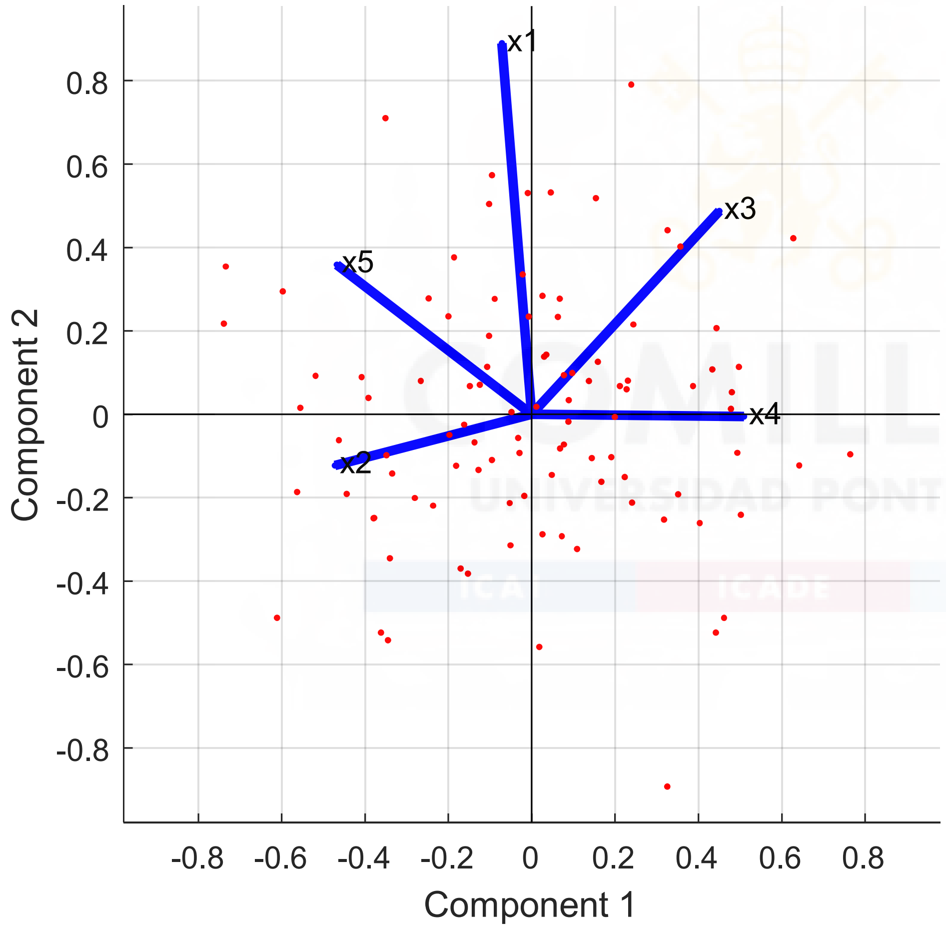
# PCA
# Illustrative synthetic cases

- C1: Uncorrelated and standardized input variables

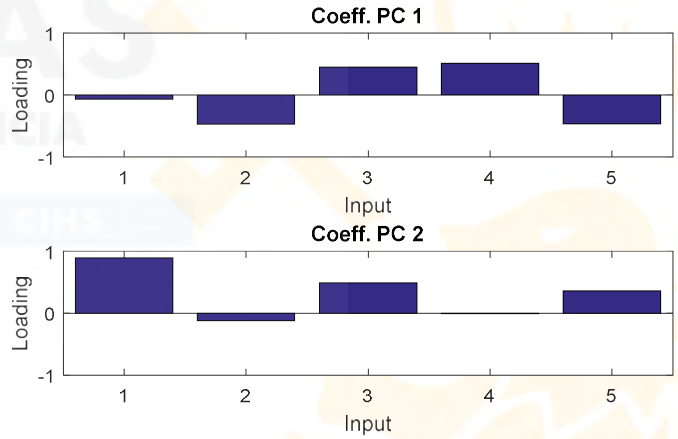


# PCA
# Illustrative synthetic cases

- C1: Uncorrelated and standardized input variables



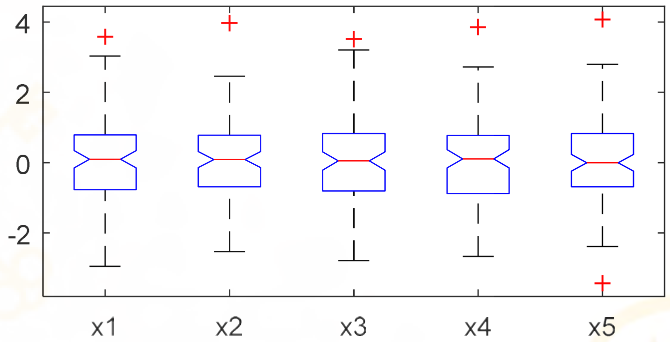

- The Biplot graph shows that there are no clear groups of input variables

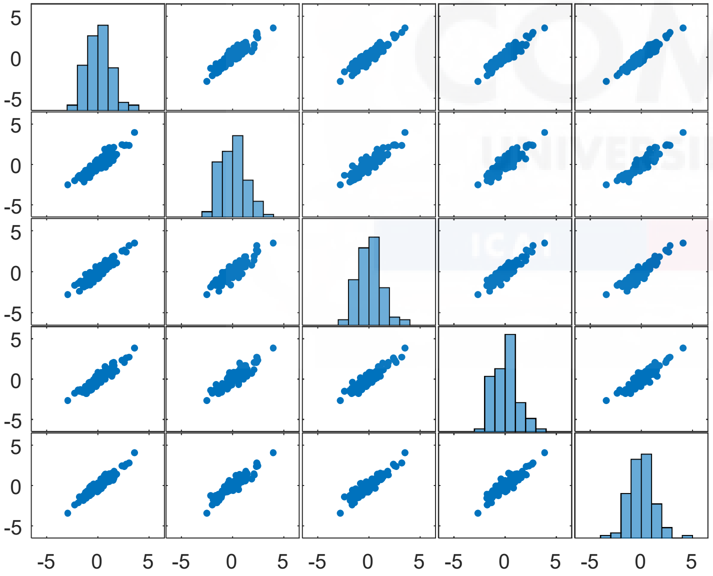


# PCA
# Illustrative synthetic cases

- C2: Very correlated and standardized input variables

```
n = 100; p = 5;
mu = zeros(p,1);
Sigma = eye(p);
Sigma(Sigma==0) = 0.9;
X=mvnrnd(mu,Sigma,n);
```




cov(X)=

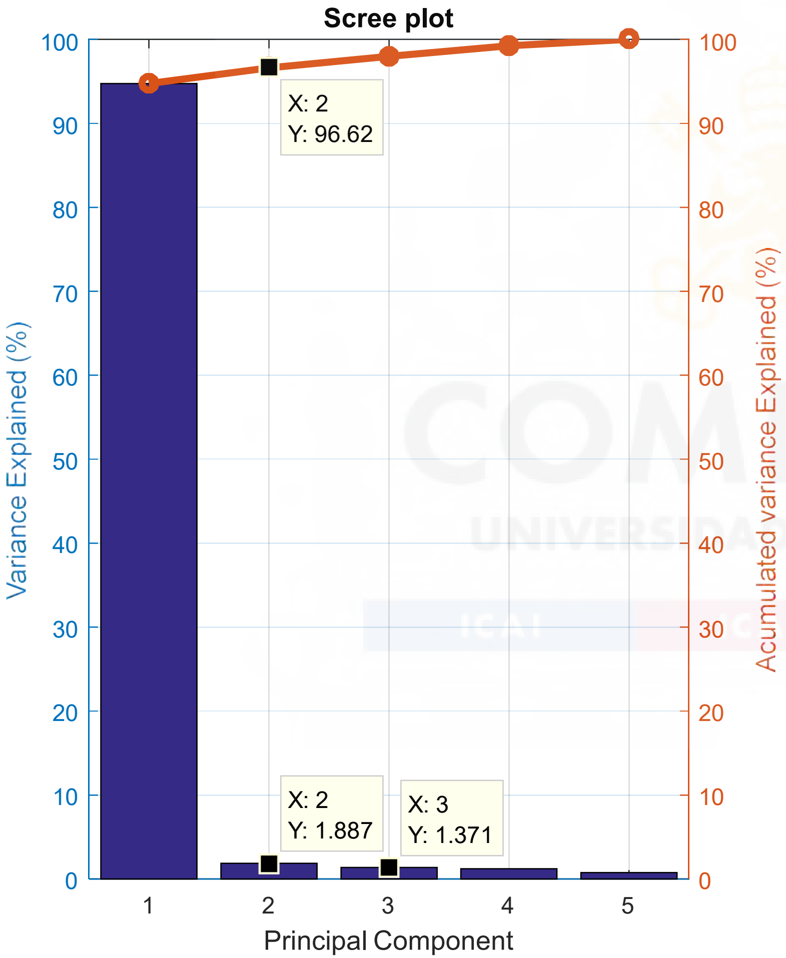
| | | | | |
|---|---|---|---|---|
| 1.3512 | 1.2545 | 1.2940 | 1.2829 | 1.3177 |
| 1.2545 | 1.3555 | 1.2588 | 1.2496 | 1.2998 |
| 1.2940 | 1.2588 | 1.4004 | 1.2928 | 1.3149 |
| 1.2829 | 1.2496 | 1.2928 | 1.3654 | 1.2871 |
| 1.3177 | 1.2998 | 1.3149 | 1.2871 | 1.4072 |

corr(X)=

| | | | | |
|---|---|---|---|---|
| 1.0000 | 0.9269 | 0.9407 | 0.9445 | 0.9557 |
| 0.9269 | 1.0000 | 0.9137 | 0.9186 | 0.9411 |
| 0.9407 | 0.9137 | 1.0000 | 0.9350 | 0.9367 |
| 0.9445 | 0.9186 | 0.9350 | 1.0000 | 0.9286 |
| 0.9557 | 0.9411 | 0.9367 | 0.9286 | 1.0000 |

# PCA
# Illustrative synthetic cases

- C2: Very correlated and standardized input variables



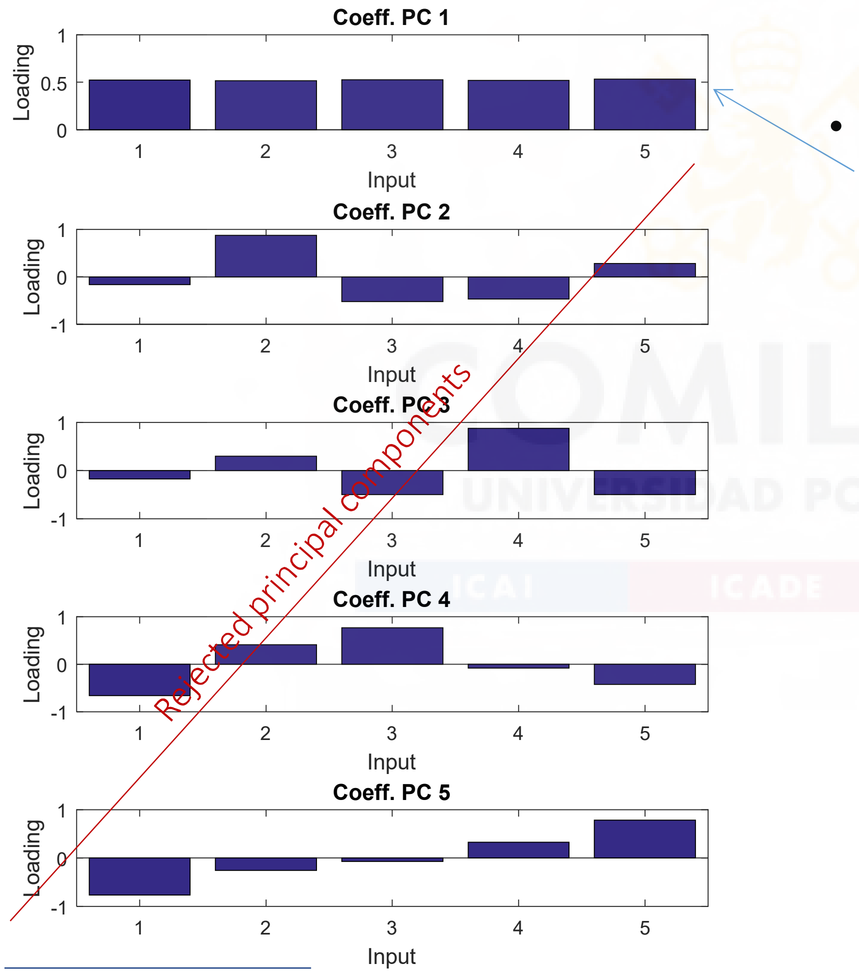

```
cov(Z)=
    4.7367   -0.0000         0   -0.0000   -0.0000
   -0.0000    0.0943   -0.0000         0   -0.0000
         0   -0.0000    0.0686    0.0000    0.0000
   -0.0000         0    0.0000    0.0616   -0.0000
   -0.0000   -0.0000    0.0000   -0.0000    0.0388
```

- The first principal component explains most of the variability in the original set $X$; it explains 94.7% of the variability
- Using the first two principal components, the accumulated proportion of variance arises to 96.62%

PCA clearly allows reducing the number of dimensions without losing much information. The first principal component summarizes the original set very well.

# PCA
# Illustrative synthetic cases

- C2: Very correlated and standardized input variables



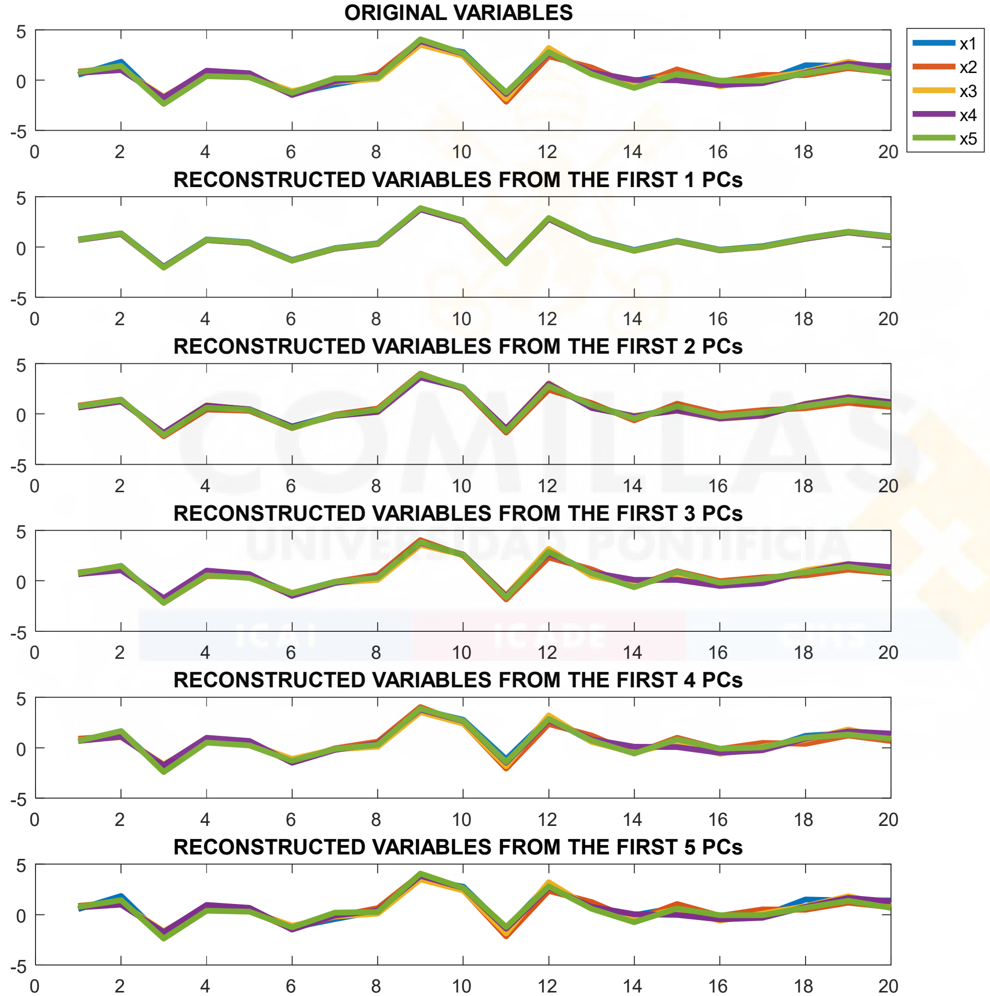

- The first principal component can be interpreted as a weighted average of all original variables with very similar weights (loadings)

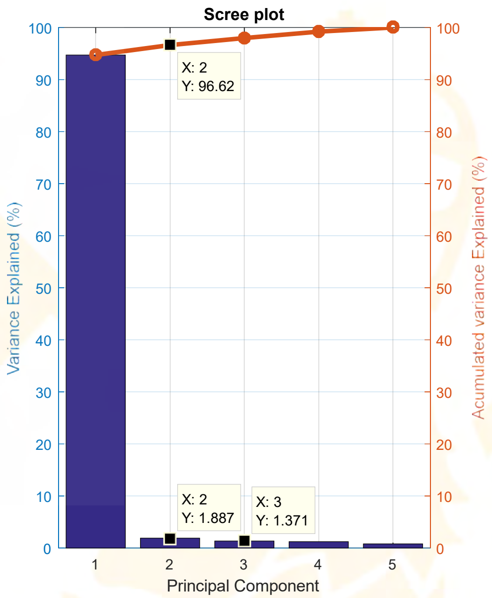
# PCA
# Illustrative synthetic cases

- C2: Very correlated and standardized input variables

94.7%

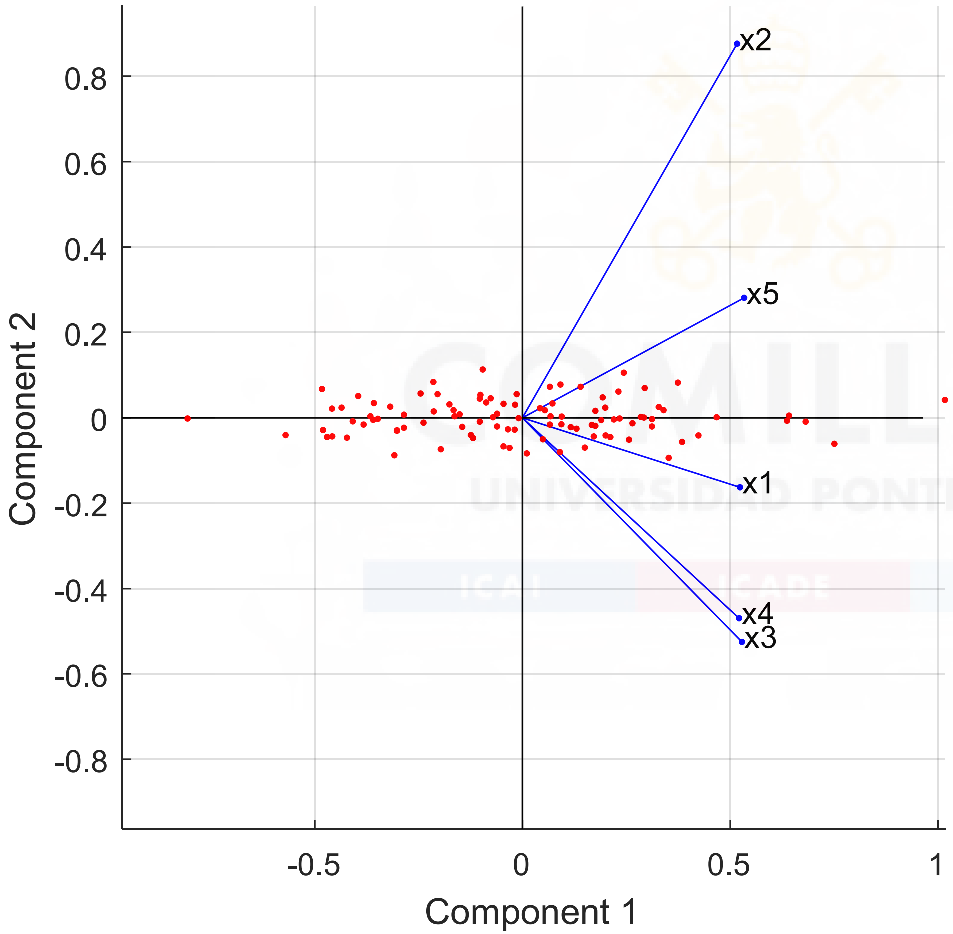
Adds only the remaining 5.3%



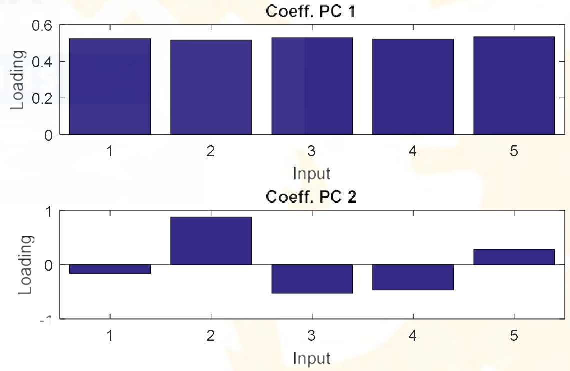



# PCA
# Illustrative synthetic cases

- C2: Very correlated and standardized input variables

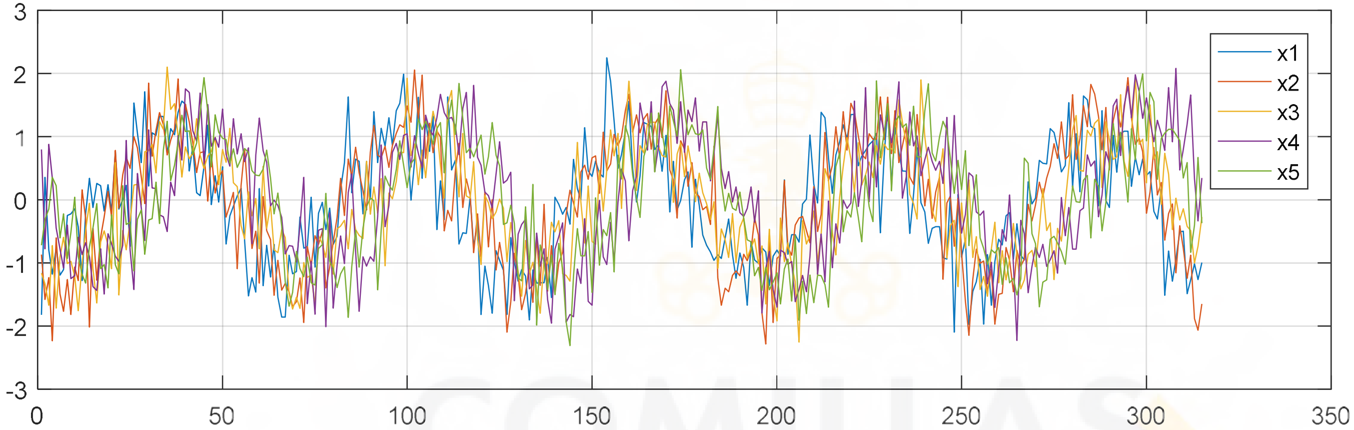


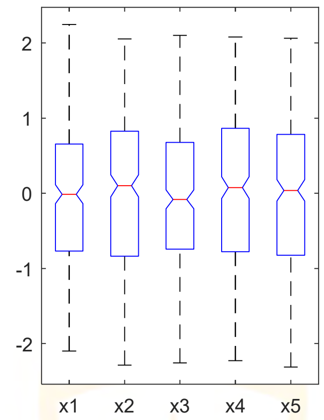

- The Biplot graph shows that the variance of the scores in the second component is very small

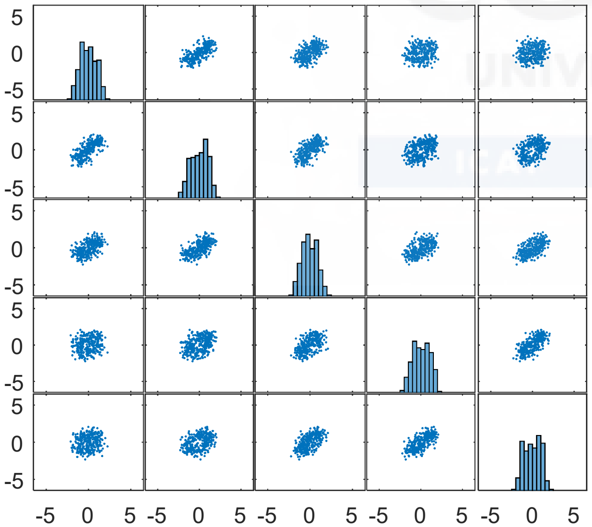


# PCA
# Illustrative synthetic cases

- C3: Mixture of correlated input variables



cov(X)=

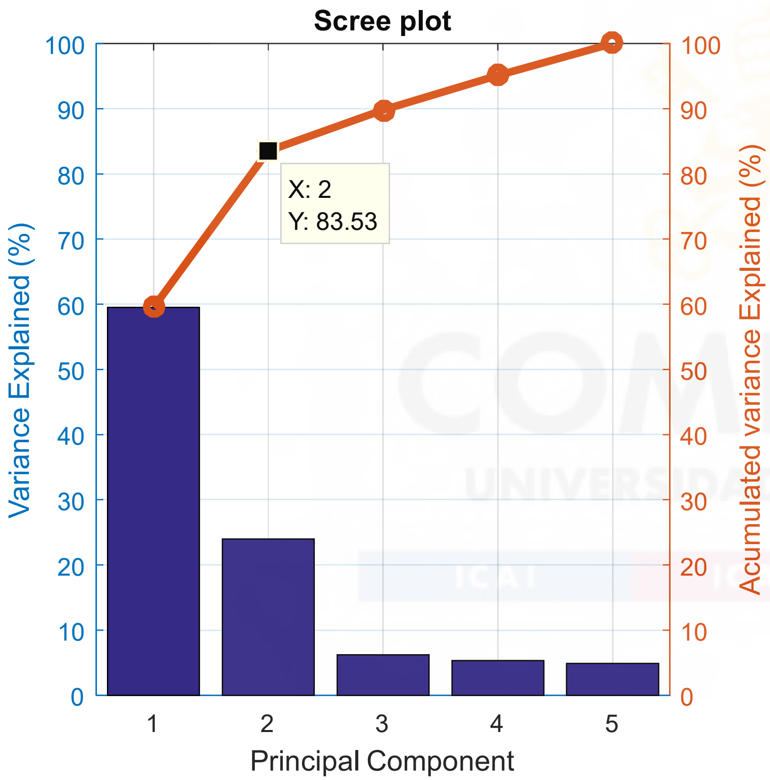
| | | | | |
|---|---|---|---|---|
| 0.8569 | 0.6774 | 0.4410 | 0.1513 | 0.1565 |
| 0.6774 | 1.0354 | 0.5988 | 0.3756 | 0.3703 |
| 0.4410 | 0.5988 | 0.7767 | 0.4832 | 0.5013 |
| 0.1513 | 0.3756 | 0.4832 | 0.9965 | 0.6922 |
| 0.1565 | 0.3703 | 0.5013 | 0.6922 | 0.9251 |

corr(X)=

| | | | | |
|---|---|---|---|---|
| 1.0000 | 0.7191 | 0.5406 | 0.1637 | 0.1757 |
| 0.7191 | 1.0000 | 0.6677 | 0.3698 | 0.3783 |
| 0.5406 | 0.6677 | 1.0000 | 0.5493 | 0.5913 |
| 0.1637 | 0.3698 | 0.5493 | 1.0000 | 0.7210 |
| 0.1757 | 0.3783 | 0.5913 | 0.7210 | 1.0000 |

# PCA
# Illustrative synthetic cases

- C3: Mixture of correlated input variables



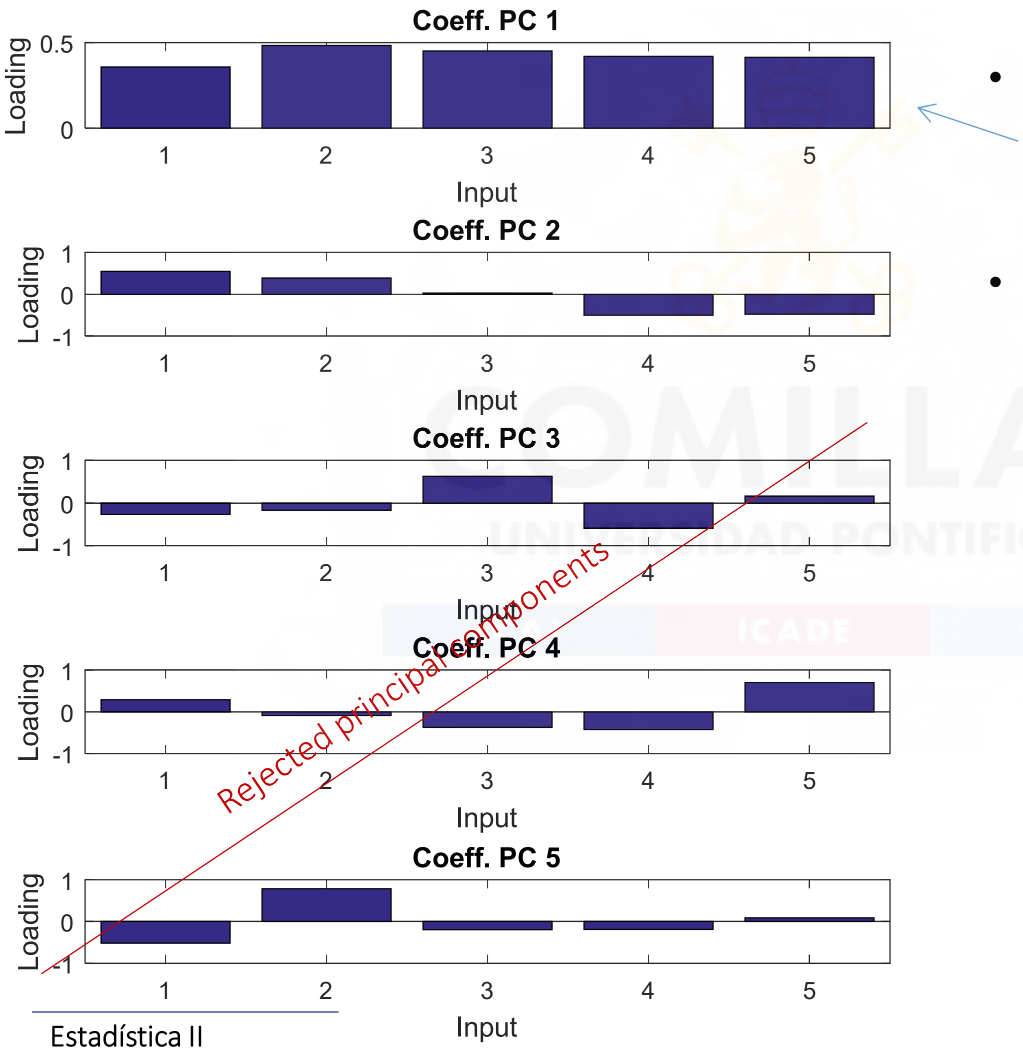

```
cov(Z)=
    2.9765    0.0000   -0.0000    0.0000    0.0000
    0.0000    1.1999    0.0000    0.0000   -0.0000
   -0.0000    0.0000    0.3119    0.0000    0.0000
    0.0000    0.0000    0.0000    0.2663    0.0000
    0.0000   -0.0000    0.0000    0.0000    0.2453
```

- The first two principal components explain most of the variability in the original set $X$. The accumulated proportion of variance of these two components is 83.5%

PCA allows for reducing the number of dimensions without losing much information. The first two principal components summarize quite well the original set.

# PCA
# Illustrative synthetic cases

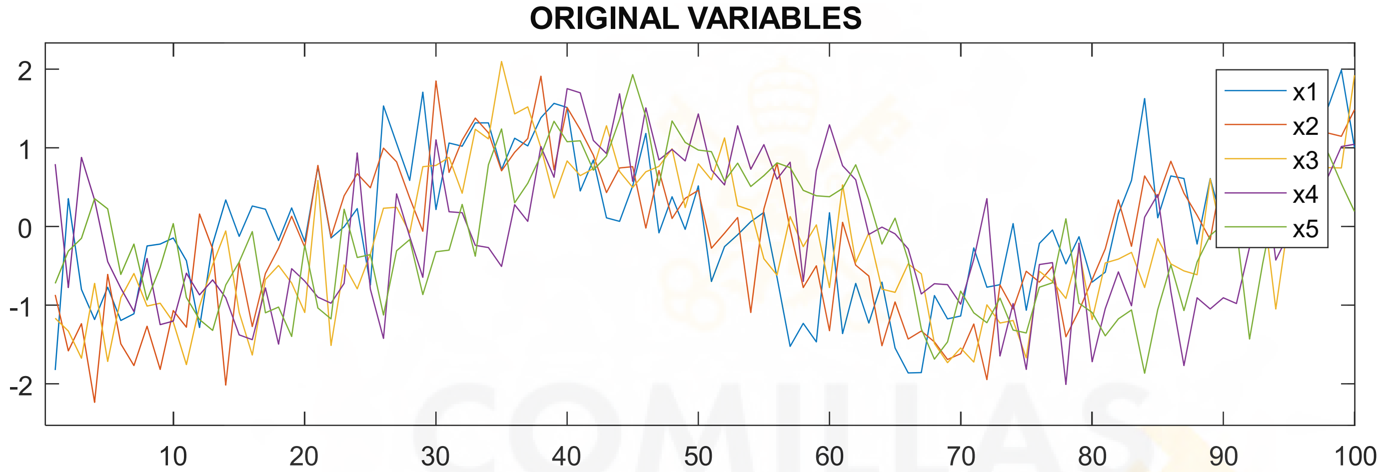
- C3: Mixture of correlated input variables

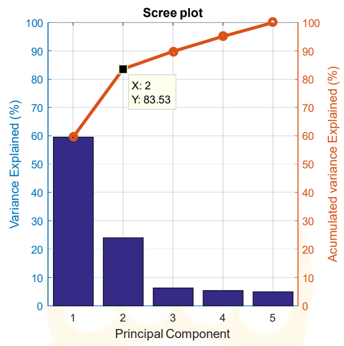


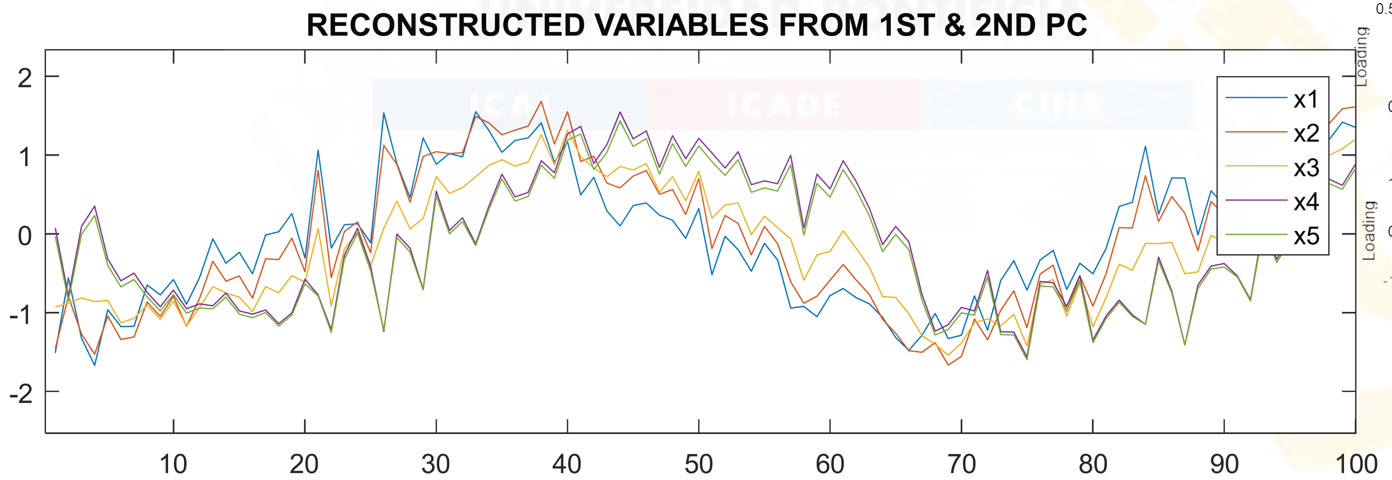

- The first principal component can be interpreted as a weighted average of all original variables with very similar weights (loadings)
- The second principal component is a shape type.

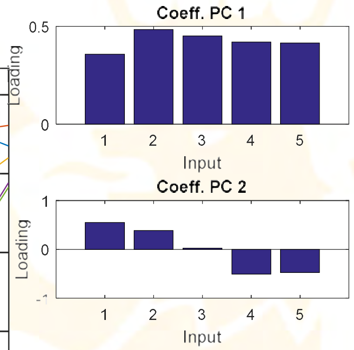
# PCA
# Illustrative synthetic cases

- C3: Mixture of correlated input variables



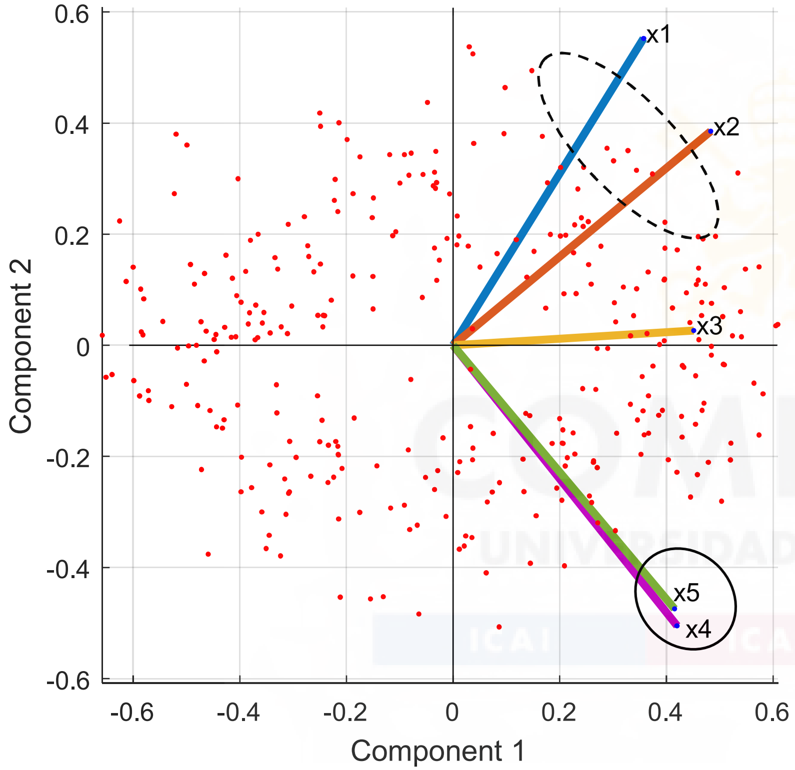



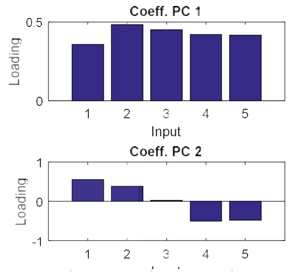



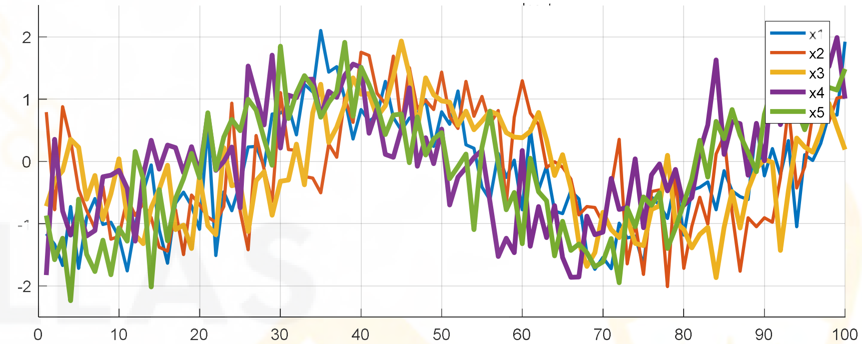



# PCA
# Illustrative synthetic cases

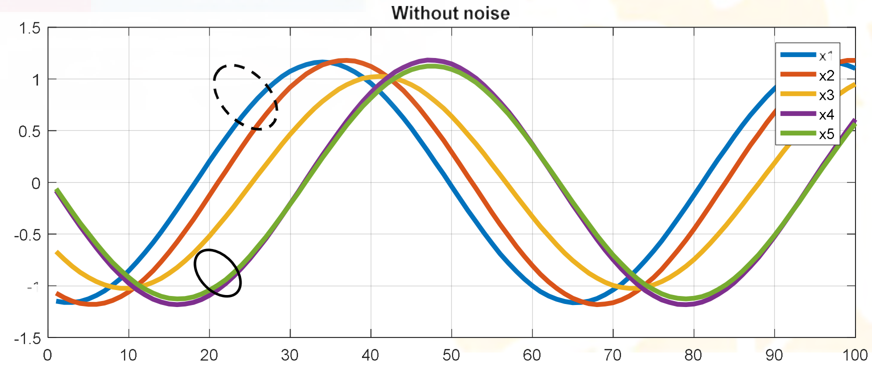
- C3: Mixture of correlated input variables




Synthetic data generated from different sine curves (adding noise)



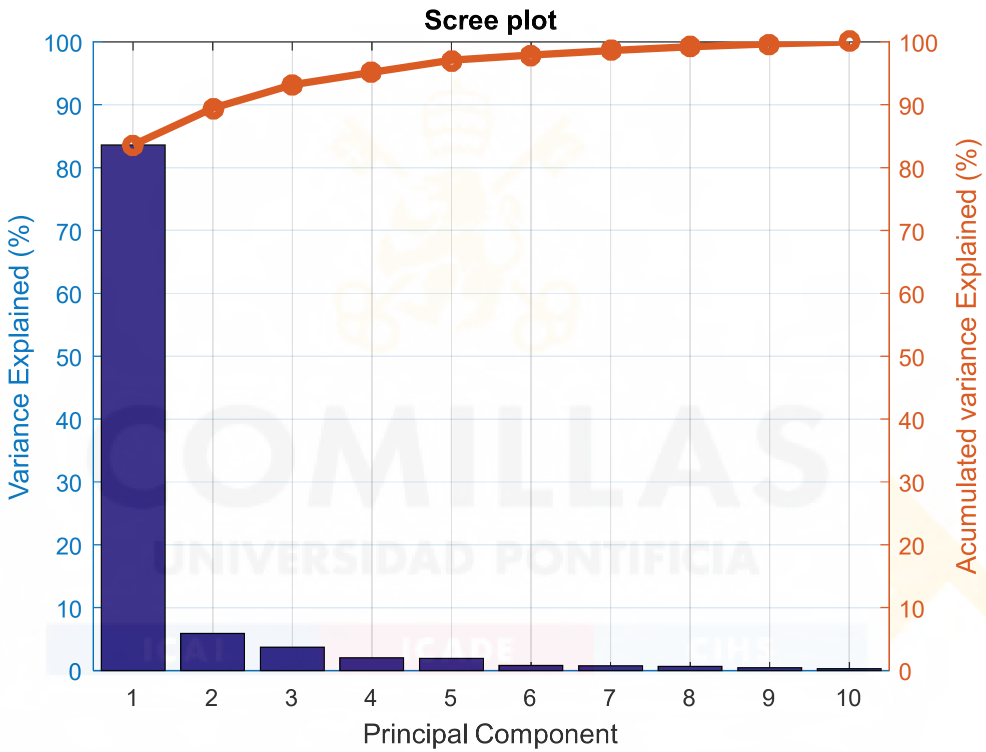

- The Biplot graph shows three groups of input variables

# 3

# Quiz

# Quiz
# Question 1

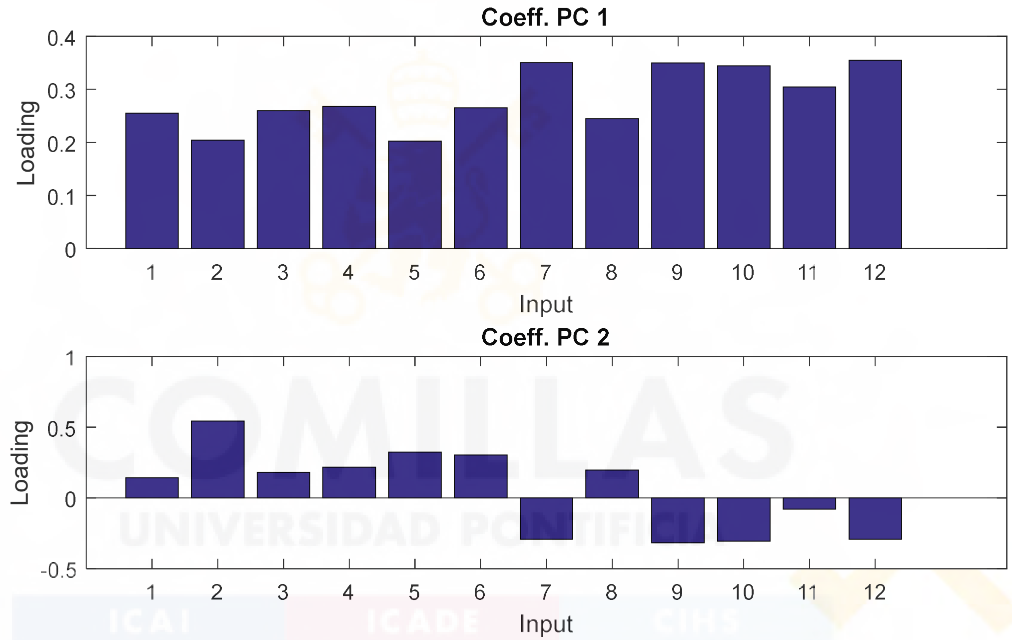
- Según el gráfico de sedimentación (scree plot) se puede afirmar que:




A. Si se eligen las dos primeras componentes principales se explica el 99.70% de la varianza total.
B. Si se eligen las cuatro primeras componentes principales se explica el 2.03% de la varianza total.
C. La tercera componente principal explica el 3.69% de la varianza total.

# Quiz
# Question 2

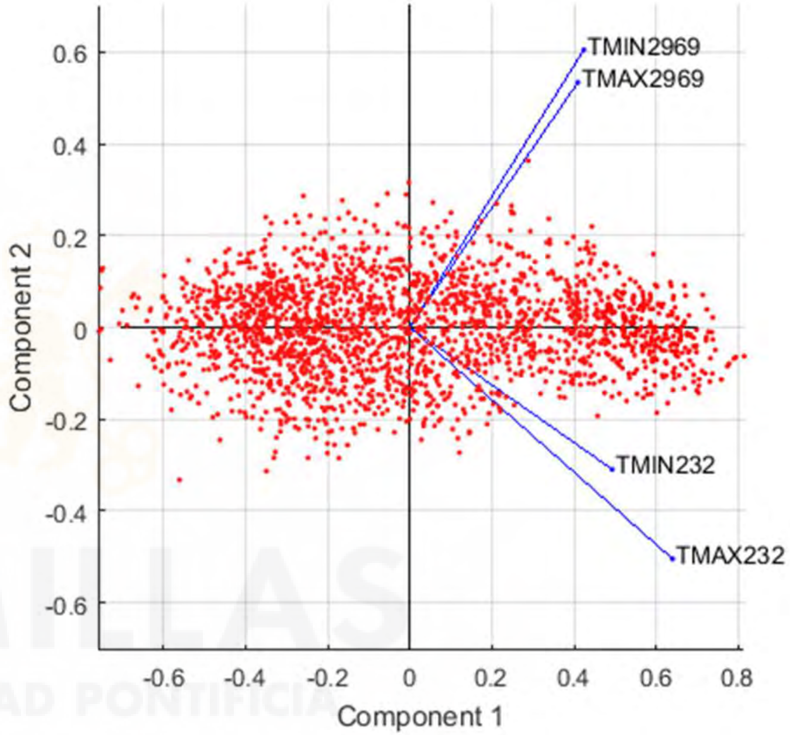
- Según los pesos que definen la primera y segunda componente principal en función de las variables originales (loadings):




A. La segunda componente principal se puede considerar como una media ponderada de todas las variables de entrada.
B. La primera componente principal se puede considerar como una media ponderada de todas las variables de entrada.
C. Tanto la primera como la segunda componente principal se pueden considerar como medias ponderadas de todas las variables, aunque cada una utiliza unos pesos diferentes. En ambas componentes se contraponen las temperaturas mínimas a las máximas.

# Quiz
# Question 3

- Según el gráfico biplot:




A. Se puede considerar que hay dos grupos claros de variables de entrada, estando uno de ellos formado únicamente por TMIN232 y TMAX232.

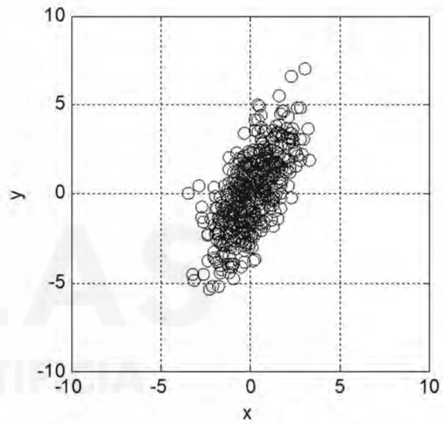
B. La primera componente se puede interpretar como un factor de forma en la que se contraponen las estaciones meteorológicas 232 y 2969.

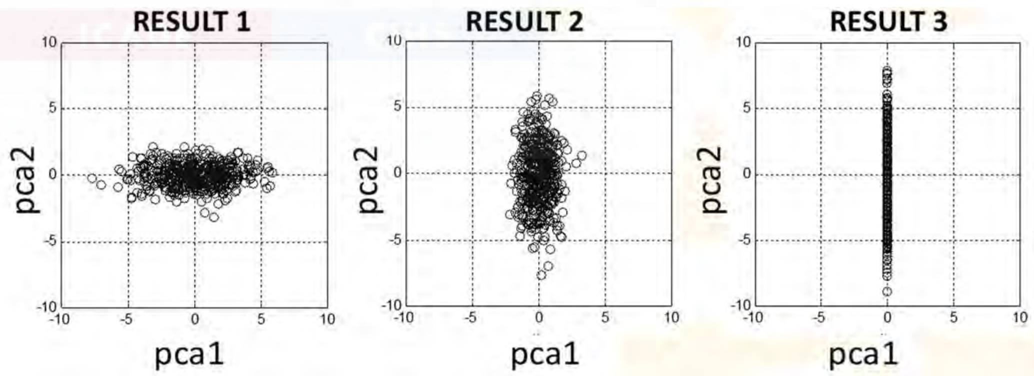
C. La primera componente se puede interpretar como un factor de forma, en la que se contraponen las temperaturas mínimas a las temperaturas máximas.

# Quiz
# Question 4

- Se tiene una muestra con 100 datos descritos por dos variables X e Y (ver figura superior). Si se realiza un análisis de componentes principales de esas variables X e Y a partir de dicha muestra, entonces se puede afirmar que al proyectar las observaciones originales en el espacio de las componentes principales dado por la primera (pca1) y segunda (pca2) componente, se obtiene:

  A. La gráfica con el título “RESULT 3”.
  B. La gráfica con el título “RESULT 1”.
  C. La gráfica con el título “RESULT 2”.



# Quiz Answers

- Q1-C
- Q2-B
- Q3-A
- Q4-B

1. Introduction
2. PCA Approach
3. Quiz
4. Real examples

4

# Real examples

# PCA real examples
# Face recognition

47th International Symposium ELMAR-2005, 08-10 June 2005, Zadar, Croatia

## Appearance-based Statistical Methods for Face Recognition

**Kresimir Delac [1], Mislav Grgic [2], Panos Liatsis [3]**
[1] Croatian Telecom, Savska 32, Zagreb, CROATIA
[2] University of Zagreb, Faculty of EE & Comp, Unska 3/XII, Zagreb, CROATIA
[3] Information and Biomedical Engineering Centre, City University London, UNITED KINGDOM
E-mail: *kdelac@ieee.org*

**Abstract** - *Different statistical methods for face recognition have been proposed in recent years. They mostly differ in the type of projection and distance measure used. The aim of this paper is to give an overview of most popular statistical subspace methods for face recognition task. Theoretical aspects of three algorithms will be considered and some reported performance evaluations will be given.*

**Keywords** - *Face Recognition, PCA, ICA, LDA, Subspace Analysis*

**1. INTRODUCTION**

Face recognition has gained much attention in recent years and has become one of the most successful applications of image analysis and

that ICA is a generalization of PCA. LDA [6], [7], unlike PCA or ICA, uses the class information and finds a set of vectors that maximizes Fisher Discriminant Criterion. It simultaneously maximizes the between-class scatter while minimizing the

# PCA real examples
# Face recognition

**2. FACE SPACE**

Generally, a two dimensional image $\Gamma(x,y)$ of size $m$-by-$n$ pixels can be viewed as a vector (or a point) in high dimensional space. The easiest way to create a vector from an array is to concatenate its columns, thus getting a vector $\boldsymbol{X} = [x_1 \ldots x_N]^T$, where $N = m \times n$. Each pixel of the image then corresponds to a coordinate in $N$-dimensional space. We will refer to this space as *image space*. Such a space has huge dimensionality ($\Re^N$) and recognition there would be computationally inefficient.

However, if an image of an object is a point in image space, a collection of $M$ images of the same sort of an object represents a set of points in the same *subspace* of the original image space. These points may be considered as samples of probability distribution. Theoretically, all possible images of one particular object define a lower-dimensional (possibly disconnected) *manifold*, embedded within the high-dimensional image space. For face recognition purposes we refer to this as *face space*. Its intrinsic dimensionality is determined by the number of degrees of freedom within face space. Appearance-based object recognition (i.e. subspace analysis) deals with the following questions [2]: what is the relationship between points in image space that correspond to all images of a particular object (face)? Is it possible to efficiently characterize this subset of all possible images? Can this subset be learned from a set of training images? What is the "shape" of this subset?

where $\boldsymbol{e}_i$ and $\lambda_i$ are eigenvectors and eigenvalues of covariation matrix $C$, respectively. We can do this because $C$ is real and symmetric. $\lambda$ is a diagonal matrix with eigenvalues on its main diagonal.

**Fig. 1.** *Mean face* calculated from 1,196 FERET gallery images [8].

Once the eigenvectors of $C$ are found, they are sorted according to their corresponding eigenvalues. Larger eigenvalue means that associated eigenvector captures more of the data variance. The efficiency of the PCA approach comes from the fact that we can

# PCA real examples
# Face recognition

Each eigenvector has the same dimensionality as a face image and looks as a sort of a "ghost" face (if rearranged and viewed as a picture), so we call them *eigenfaces* (Fig. 5, top row). Transforming a point to

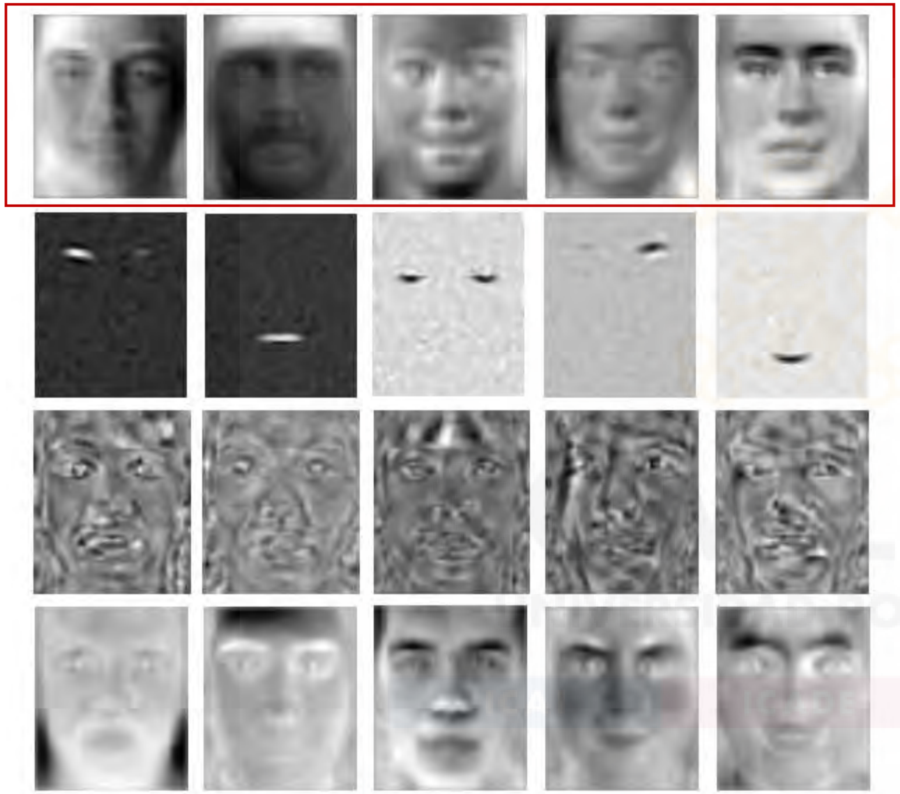


**Fig. 5.** An illustrative example of differences between PCA, ICA *Architecture I*, ICA *Architecture II* and LDA. Top row shows top eight PCA eigenfaces. The second row shows localized feature vectors for ICA *Architecture I*. The third row shows eight non-localized ICA feature vectors for ICA *Architecture II*. Bottom row shows LDA representation vectors (*Fisherfaces*).

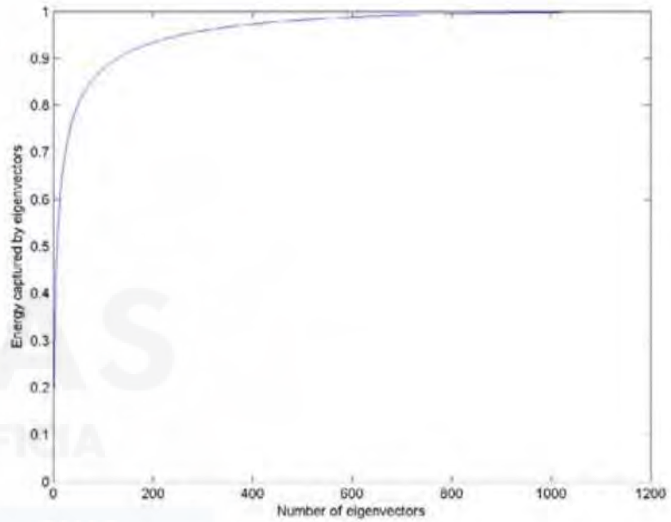



**Fig. 2.** Energy captured by retaining the number of eigenvectors with the largest eigenvalues, calculated from 1,196 FERET gallery images [8]. It is clearly seen that retaining only 200 eigenvectors (of total 1,196 vectors) captures more then 90% of the energy.

# PCA real examples
# Residual demand on electricity markets

50 IEEE TRANSACTIONS ON POWER SYSTEMS, VOL. 32, NO. 1, JANUARY 2017

## Residual Demand Curves for Modeling the Effect of Complex Offering Conditions on Day-Ahead Electricity Markets

José Portela González, Antonio Muñoz San Roque, Eugenio F. Sánchez-Úbeda, Javier García-González, and Rafael González Hombrados

***Abstract*—Residual demand curves (RDCs) can be used to represent the strategic interaction of participants in electricity markets. RDCs relate the energy that an agent can buy or sell in one hour with the clearing market price that would be obtained in such hour, assuming the market is organized as simple bid independent auctions. Despite the fact that they have been widely used in the literature, the existence of time and/or spatial constraints in the market clearing algorithm makes the RDCs not directly applicable. This paper tries to overcome these difficulties by extending the concept of RDCs to zonal pricing markets where complex offering conditions and transmission constraints are taken into account. Therefore, the RDCs are redefined in order to capture such effects, which are usually neglected or oversimplified. A new method for computing the redefined RDCs is established and its application to the Iberian electricity market is presented. The results show that modeling complex conditions and transmission constraints in RDCs can have a significant effect when compared to the standard approach found in the literature. Therefore, the method presented in this paper modeling the effect of firm's decisions on market prices in a more accurate way.**

***Index Terms*—Complex offers, electricity markets, inter-temporal constraints, market splitting, residual demand curves, transmission constraints.**

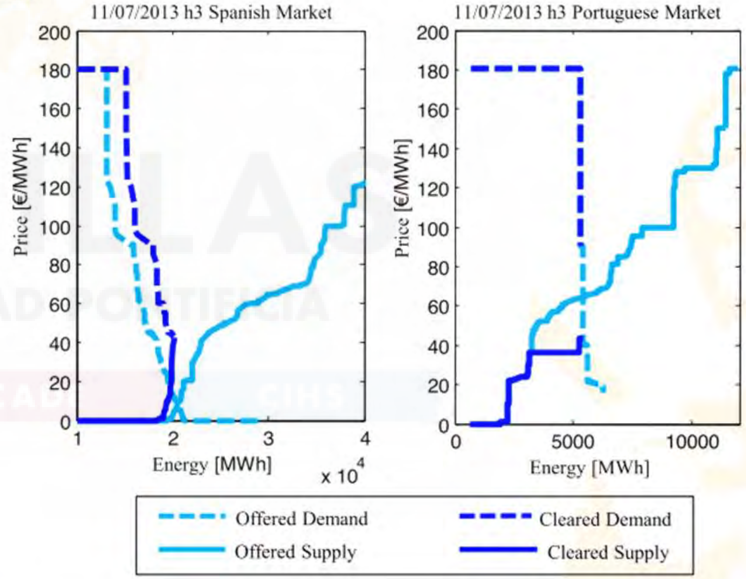



Fig. 2. Aggregated supply and demand curves for Spain and Portugal when Market Splitting has occurred (11/07/2013, hour 3). In this case, the congestion of the transmission capacity has occurred when the energy is being exported from Spain to Portugal. Then, the Spanish offered demand is only showing the demand bids of Spanish agents while the Spanish cleared demand includes a demand bid with the energy value of the interconnection capacity to model the selling to the Portuguese market.

# PCA real examples
# Residual demand on electricity markets

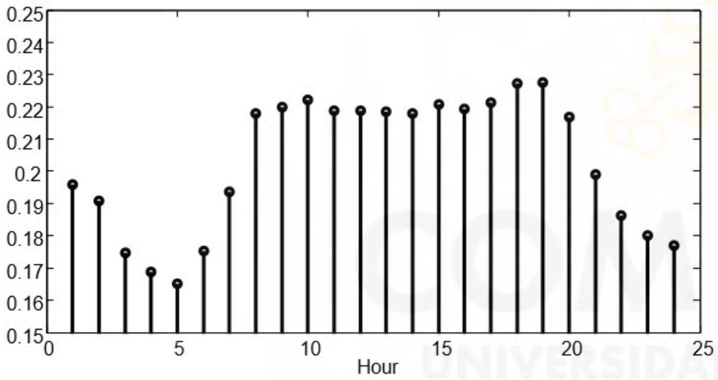



Figure 5: Values of the first principal component extracted for daily price of 2013. Higher values mean that prices change at a higher rate than hours with lower values.

In order to build the rest of the curve, it is necessary to sample the 24-dimensional price space. Historically it can be observed that the daily price profiles are related, evidencing that the embedded dimension is smaller. The dimensionality can be reduced through Principal Component Analysis (PCA) [25]. This decomposition technique finds the linear transformation that forms a subspace that maximizes the variance of the data. The first principal component extracted $\boldsymbol{v} = (v_1, \dots, v_{24})$ (the eigenvector associated to the highest eigenvalue of the transformation matrix) is the direction that models the highest percentage of the variance of the data. Figure 5 plots the eigenvector $\boldsymbol{v}$ obtained when PCA is applied to the daily price profiles of year 2013 in the Spanish market, accounting for the 82.81 percent of the variance of the data. Consequently, we decide to reduce the price space to one dimension.

It is important to highlight that the temporal link that exists in all the hours is being modeled thanks to this PCA reduction.

Then, in order to create a full curve, the 24 dimensional space of price profiles is explored by searching along the embedded dimension given by $\boldsymbol{v}$. In the appendix, a sensitivity study shows the results when different search paths are used.

A

# Dictionary

PCA

# Dictionary

- **Bimodal** – bimodal (dos modas)
- **Biplot** – gráfico doble
- **Eigenvector** – autovector
- **Eigenvalue** – autovalor
- **Elbow** – codo (punto de inflexión en una curva)
- **Pattern** – patrón, perfil
- **Principal Component Analysis (PCA)** – Análisis de componentes principales
- **Scree plot** – gráfico de sedimentación

*Thank you for your attention*

Eugenio Sánchez Úbeda