UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial,
Aeroespacial i Audiovisual de Terrassa

MUEI, MUEA, and MASE

Advanced Engineering Data Analysis (AEDA)
Principal Components Analysis
Lab Practice

Prof. Daniel Fernández
daniel.fernandez.martinez@upc.edu

Lab practice 1

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

Understanding USArrests data using PCA

Description: This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973.

Also given is the percent of the population living in urban areas.

A data frame with 50 observations on 4 variables.

- **Murder** (numeric): Murder arrests (per 100,000)
- **Assault** (numeric): Assault arrests (per 100,000)
- **UrbanPop** (numeric): Percent urban population
- **Rape** (numeric): Rape arrests (per 100,000)

More details: https://www.rdocumentation.org/packages/datasets/versions/3.6.2/topics/USArrests

Activity: Apply a pre-processing of the data and apply PCA. Contextualize the results. Follow the example showing in class and explore PCA aspects by yourself.

Solution proposal: https://rstudio-pubs-static.s3.amazonaws.com/377338_75ed92a8463d482a80045abcae0e395d.html

Lab practice 2

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

US Air pollution using PCA

Description: Air pollution data of 41 US cities

Air pollution data of 41 US cities

	SO2	temp	manu	popul	wind	precip	predays
Albany	46	47.6	44	116	8.8	33.36	135
Albuquerque	11	56.8	46	244	8.9	7.77	58
Atlanta	24	61.5	368	497	9.1	48.34	115
Baltimore	47	55.0	625	905	9.6	41.31	111
Buffalo	11	47.1	391	463	12.4	36.11	166
Charleston	31	55.2	35	71	6.5	40.75	148

More details: https://www.rdocumentation.org/packages/HSAUR3/versions/1.0-10/topics/USairpollution

Location:

```
library("HSAUR2")
data(USairpollution)
```

Activity: Apply a pre-processing of the data and apply PCA. Contextualize the results. Follow the example showing in class and explore PCA aspects by yourself.

Solution proposal: USairpollution.html on Atenea

Lab practice 3

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

Iris data set using PCA

Description: It gives the measurements in centimetres of the variables: sepal length and width and petal length and width for 50 flowers from each of 3 species of iris: setosa, versicolor, and virginica.

Edgar Anderson's Iris Data

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

More details:
https://www.rdocumentation.org/packages/datasets/versions/3.6.2/topics/iris

Location:

```
library(”datasets")
data(iris)
```

Activity: Apply a pre-processing of the data and apply PCA. Contextualize the results. Follow the example showing in class and explore PCA aspects by yourself.

Solution proposal: https://rpubs.com/amos593/419546

Lab practice 4

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

Athletes' performance in decathlon using PCA

Description: Athletes' performance during two sporting meetings. It contains 27 individuals (athletes) described by 13 variables

We will be using a subset of the data set.

More details: https://www.rdocumentation.org/packages/factoextra/versions/1.0.7/topics/decathlon2

Location:

```
library("factoextra")
data(decathlon2)
decathlon2.active <- decathlon2[1:23, 1:10]
```

Activity: Apply a pre-processing of the data and apply PCA. Contextualize the results. Follow the example showing in class and explore PCA aspects by yourself.

Solution proposal: http://www.sthda.com/english/articles/31-principal-component-methods-in-r-practical-guide/112-pca-principal-component-analysis-essentials/