Symbolic Formulation for Principal Component Analysis of Interval-Valued Data

Rodrigo Girão Serrão

rodrigogiraoserrao@gmail.com

Winter School in Mathematics 2021

February 24, 2021

supervised by
Professor Lina Oliveira
Professor M. Rosário Oliveira

with the support of
Instituto de Telecomunicações

Overview

Analyse grades of students.

Analyse grades of students.

Conventional approach:

- average
- variance
- minimum and maximum
- ...

Analyse grades of students.

Conventional approach:

- average
- variance
- minimum and maximum
- ...

Single valued variables summarise object of interest.

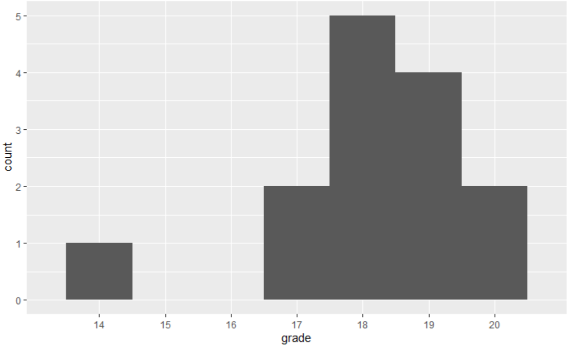
Analyse grades of students.

Symbolic approach:

Analyse grades of students.

Symbolic approach:

- histogram



Analyse grades of students.

Symbolic approach:

- histogram
- interval: $[14, 20]$
- ...

Analyse grades of students.

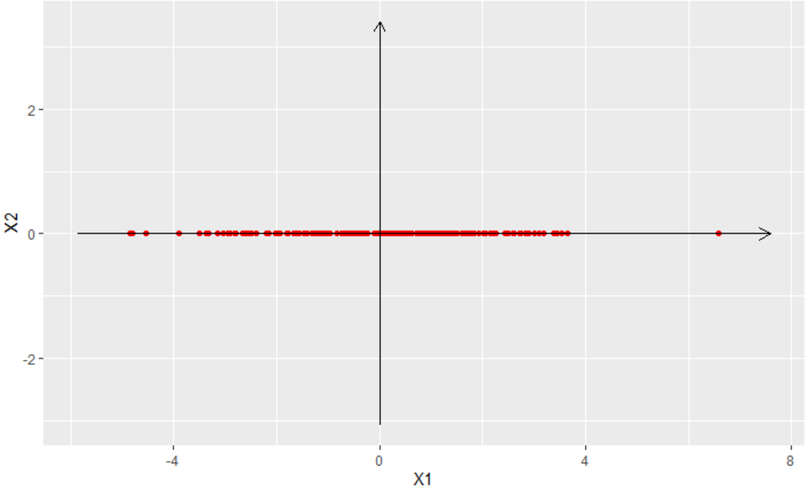
Symbolic approach:

- histogram
- interval
- ...

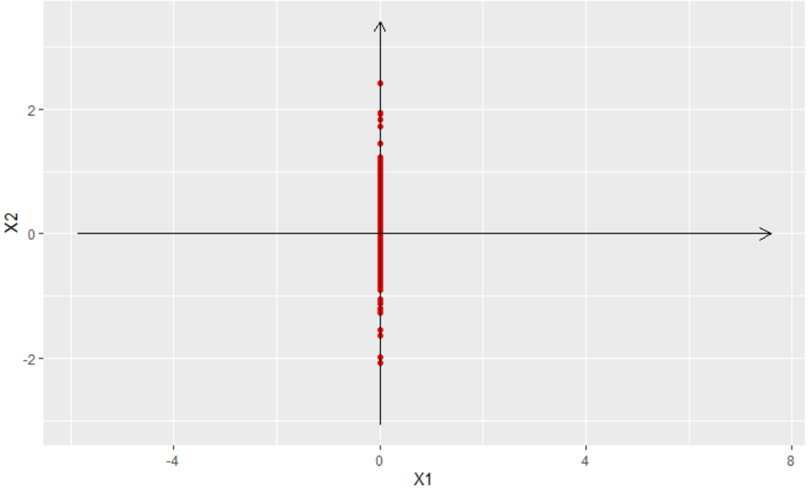
Data account for intrinsic variability.

Let us study some 2d data.

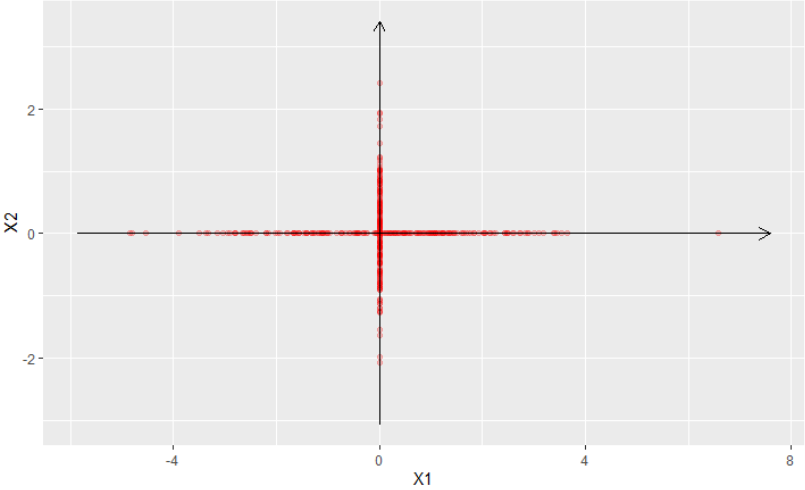
Let us study some 2d data... along the x axis.



Let us study some 2d data... along the y axis.

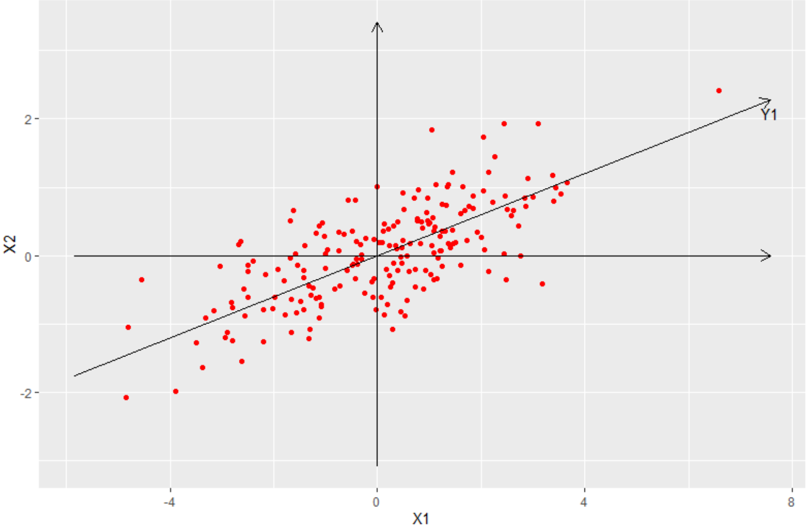


Let us study some 2d data... which one is best?

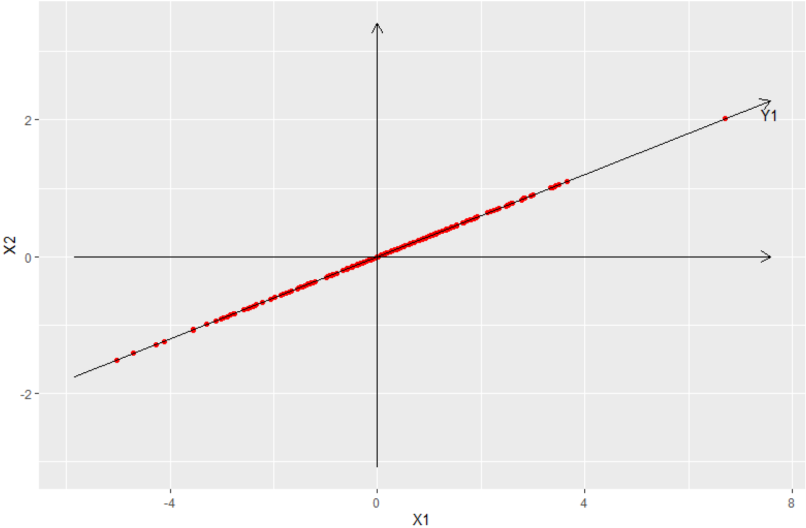


Is that the best we can do?

Is that the best we can do? – No



Is that the best we can do? – No



How to pick that direction Y_1?

How to pick that direction Y_1?
Write $Y_1 = \gamma_1^T \boldsymbol{X}$,

How to pick that direction Y_1?
Write $Y_1 = \gamma_1^T \boldsymbol{X}$,

$$\gamma_1 = \arg\max_{\gamma : ||\gamma|| = 1} \mathrm{Var}(\gamma^T \boldsymbol{X})$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{PCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{PCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{PCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma : ||\gamma|| = 1} \mathrm{Var}(\gamma^T \boldsymbol{X}) \\ \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{PCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma : ||\gamma||=1} \mathrm{Var}(\gamma^T \boldsymbol{X}) \\ \mathrm{Cov}(\gamma^T \boldsymbol{X}, \gamma_j^T \boldsymbol{X}) = 0,\ j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{PCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma : ||\gamma|| = 1} \mathrm{Var}(\gamma^T \boldsymbol{X}) \\ \gamma^T \gamma_j = 0, \; j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{SPCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma:||\gamma||=1} \mathrm{Var}(\gamma^T \boldsymbol{X}) \\ \gamma^T \gamma_j = 0, \ j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{SPCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma : ||\gamma|| = 1} \mathrm{Var}(\gamma^T \boldsymbol{X}) \\ \gamma^T \gamma_j = 0, \; j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{SPCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma : ||\gamma|| = 1} \mathrm{Var}(\gamma^T \boldsymbol{X}) \\ \gamma^T \gamma_j = 0, \; j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{SPCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma:||\gamma||=1} \text{Var}(\gamma^T \boldsymbol{X}) \\ \gamma^T \gamma_j = 0, \ j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Random vector $\boldsymbol{X} = (X_1, \cdots, X_p)^T \xrightarrow{\text{SPCA}} \boldsymbol{Y} = (Y_1, \cdots, Y_p)^T$, vector of the principal components.

$Y_i = \gamma_i^T \boldsymbol{X}$,

$$\gamma_i = \begin{cases} \arg\max_{\gamma:||\gamma||=1} \text{Var}_k(\gamma^T \boldsymbol{X}) \\ \gamma^T \gamma_j = 0, \ j = 1, \cdots, i-1, \quad \text{if} \quad i > 1 \end{cases}$$

Some other tasks:

- Implement the method;
- Create a robust variant;
- Explore real data;

Thank you.