1- Description

This repository contains code to reproduce results in

L. Forzani, R. Garcia Arancibia, P. Llop and D. Tomassi, "*Sufficient dimension reduction for ordinal predictors*" (submitted).

The main code for the proposed method is written in Matlab. Nevertheless, some comparisons with other methods require running scripts in R. Indications to reproduce results reported in the manuscript are detailed below.

2- Organization

The code is organized in several folders:

./SCRIPTS: here are the scripts to reproduce the results reported in the manuscript.

./MAINFUNS: here are the functions which are called to apply the proposed methods.

./INTERNALS: here are the procedures and auxiliary functions to actually implement the proposed methods.

./TOOLKITS: here are other tools we take advantage of when implementing and running the code.

3- Usage

To start using the code, set Matlab's working directory to the main folder of this package and run

```
>setpaths
```

This will add paths to all the internal folders so that all the functions and datasets become available.

Reproducing experiments with synthetic data

To reproduce Figure 1, run

```
>simulation_figure1a
>simulation_figure1b
```

To reproduce Figure 2, run

```
>simulation_figure2
```

To reproduce Table 1, run

```
>Inference_perm
>Inference_cv
```

To reproduce Table 2, run

```
>simulation_Table2
```

Reproducing experiments with real data

In section 6.1 of the manuscript we apply the proposed method to build an index of socio-economic status. The data can be found in the folder ./SCRIPTS/DATA/EPH.

To reproduce results reported in Table 3 you need to run some scripts in Matlab and some in R.

In Matlab, run

```
>Tabla3_orig1_continuous
>Tabla3_orig1_binary
>Tabla3_PFCord_continuous
>Tabla3_PFCord_binary
```

In R, run

```
>source("cv.ordSelect.R")
>source("Tabla3_continuousORIG2.R")
>source("Tabla3_binaryORIG2.R")
>source("Tabla3_continuousLASSOord.R")
>source("Tabla3_binaryLASSOord.R")

>source("cv.PCApoly.R")
>source("Tabla3_continuousPCApoly.R")
>source("Tabla3_binaryPCApoly.R")
```

To reproduce results reported in Table 4 you need to run some scripts in Matlab and some in R.

In Matlab, run

```
>loadingsEPH_PCApoly
>loadingsEPH_PFCord_continuous
>loadingsEPH_PFCord_binary
```

In R, run

```
>source('auxfuns.R')
>source('loadingsEPH_PCApoly.R')
>source('loadingsEPH_NLPCA.R')
```

In section 6.2 of the manuscript we apply the proposed method to a movie recommendation problema. The data can be found in the folder ./SCRIPTS/DATA/Netflix. To reproduce results reported in Table 6 run

```
>script_netflix
```

Please, be aware that the script can take several weeks to complete. Furthermore, you need to install the **libsvm** library if you want to reproduce the full set of results. Find instructions in the script.