NIPALS_PCA

A Julia package for calculating PCA and PLS using the NIPALS implementation. Both models handles missing values

The package contains data structures for models and datasets

Installation

In Julia add https://gitlab.moffitt.usf.edu:8000/Bios2Projects/NIPALS_PCA as unregistered package

```
using Pkg
Pkg.add("https://gitlab.moffitt.usf.edu:8000/Bios2Projects/NIPALS_PCA")
```

Running Julia REPL

Julia can be started using

1. The base installation
2. From Singularity container (in progress, details pending)
3. Utilizing the downloaded folder as a local environment.

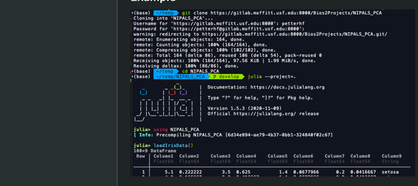
To activate NIPALS_PCA as local environment

```
cd path/to/cloned/NIPALS_PCA
julia --project=.
```

Loading package

```
using NIPALS_PCA
```

Example



On cluster using Singularity

A prebundled singularity container for Julia can be accessed at /share/data2/applications/singularity_images/bbsrTools.sif

-C, is required to utilize packages installed in the container. If this flag is not used, Julia will load packages from the users home directory

-B, binds folders to be accessible to the container

```
module load singularity/3.10
sudo singularity run -C -B ../data:/data  bbsrTools.sif
```

Tutorial

PCA modelling

From Julia REPL

1)Load package

```
using NIPALS_PCA
```

2)Load dataset from .csv file to DataFrame

```
x_df = loadIrisData()
```

3)Create dataset and apply normalize to mean center data

```
xdataset = parseDataFrame(x_df) |> normalize
```

4)Calculate PCA model

```
pca = calcPCA(xdataset, 3)
```

5)Calculate variances for model

```
calcVariances(xdataset,pca)
```

PLS normalization

The PLS normalization workflow can either be run from a script or from an interactive Julia session. The default script can be find in src/scripts/plsnorm.jl

Run from script

```
julia src/scripts/plsnorm.jl \
--xfile /path/to/xmatrix.txt \
--yfile /path/to/ymatrix.txt \
--ycategorical "colname" \
--ycontinous "colname1:colname2" \
--mode calibrate \
--modelfile model.jld2 \
--outfile output_file.csv
```

Run from interactive session

```
using NIPALS_PCA

parsed_args=Dict{String,Any}("xfile" => "/path/to/xmatrix.txt","ycategorical" => "colname","yfi

#to calibrate
calibrate_model(parsed_args)

#to correct
correct(parsed_args)
```

Run on cluster using Singularity image

The folder(s) used for reading and writing needs a Bind. In the example below
 -B ../data/:/data
 data will be accessible from the root level within the container. Multiple folders can be bound.

```
module load singularity/3.10

sudo singularity run --app plsnorm  -C -B ../data/:/data bbsrTools.sif \
--xfile /data/xmatrix.txt \
--yfile /data/ymatrix.txt \
--ycategorical 'colname' \
--ycontinous 'colname1:colname2' \
--mode calibrate \
--modelfile model.jld2 \
--outfile output_file.csv
```

Get help

```
julia plsnorm.jl --help
```

Structures

Functions

General functionality

PLS normalization