LIBRA: a MATLAB Library for Robust Analysis

List of Functions

June 28, 2016

This document contains the list of functions that are currently available in the 'MATLAB Library for Robust Analysis'. This toolbox is developed at ROBUST@Leuven, the research group on robust statistics at the KU Leuven and can be downloaded from the website

http://wis.kuleuven.be/stat/robust/LIBRA

It contains user-friendly implementations of many robust procedures, most of them being developed at ROBUST@Leuven. These methods are resistant to outliers in the data. Many graphical tools are provided for model checking and outlier detection. Many functions require the MATLAB Statistics Toolbox.

Contributions to this toolbox have been made by (in alphabetical order): Guy Brys, Michiel Debruyne, Sanne Engelen, Mia Hubert, Wai Yan Kong, Nele Smets, Wannes Van den Bossche, Karlien Vanden Branden, Stephan Van der Veeken, Ellen Vandervieren, Katrien Van Driessen, Sabine Verboven, Tim Verdonck and Fabienne Verwerft.

The toolbox can be freely used for non-commercial use only. Please make appropriate references to the corresponding paper(s) if you use any of our programs. The correct references can be found in the help-files, or at the web page:

http://wis.kuleuven.be/stat/robust

More details on the use of the library are described in:

Verboven, S., Hubert, M. (2005), LIBRA: a MATLAB Library for Robust Analysis, *Chemometrics and Intelligent Laboratory Systems,* 75, 127-136.

Verboven, S., Hubert, M. (2010). Matlab library LIBRA, *Wiley Interdisciplinary Reviews: Computational Statistics,* 2, 509-515.

Bugs or comments on the programs can be reported to Mia Hubert:

Mia.Hubert@kuleuven.be

Name	Description	Available since
Robust estimators of location, scale, skewness.		
mlochuber	M-estimator of location with Huber psi-function	22-04-2004
mloclogist	M-estimator of location with logistic psi-function	22-04-2004
hl	Hodge-Lehmann location estimator	22-04-2004
unimcd	MCD estimator of location and scale	30-06-2003
mad	Median absolute deviation	30-06-2003
mscalelogist	M-estimator of scale with logistic psi-function	22-04-2004
qn	Qn-estimator of scale	30-06-2003
adm	Scale estimator given by the Average Distance to the Median	22-04-2004
mc	Medcouple: robust estimator of skewness	22-04-2004
robstd	Columnwise robust standardization	22-04-2004
adjustedboxplot	Computes and plots skewness-adjusted boxplot	22-12-2006
Robust multivariate analysis.		
l1median	L1-median of multivariate location	30-06-2003
mcdcov	Minimum Covariance Determinant estimator of multivariate location and covariance computed using the FAST-MCD algorithm	22-04-2003
DetMCD	MCD estimator computed using the faster but not fully equivariant DetMCD algorithm	31-10-2011
ddc	Detect deviating cells	26-06-2016
cellmap	Plots the results of ddc as a heatmap	26-06-2016
rapca	Robust principal component analysis (based on projection pursuit)	30-06-2003
robpca	Robust principal component analysis (based on projection pursuit and MCD estimation)	30-06-2003
rda	Robust linear and quadratic discriminant analysis (classification of low-dimensional data)	22-04-2004
classifskew	Robust classification of low-dim skewed data	31-10-2011
rsimca	Robust soft independent modelling of class analogies (classification of high-dimensional data)	20-09-2004
adjustedoutlyingness	Detection of multivariate outliers at skewed data: based on the adjusted outlyingness at symmetric data: based on the Stahel-Donoho outlyingness	25-02-2008
halfspacedepth	Halfspace depth of bivariate data points	25-02-2008
bagplot	Draws the bagplot of bivariate data points, based on halfspace depth or adjusted outlyingness. Also yields the Tukey median (deepest point) and the halfspace depth of all observations.	25-02-2008

Name	Description	Available since
Robust regression methods.		
ltsregres	Least Trimmed Squares regression	30-06-2003
mcdregres	Multivariate MCD regression	30-06-2003
rpcr	Robust principal component regression	30-06-2003
rsimpls	Robust partial least squares regression	30-06-2003
cdq	Censored depth quantiles	26-07-2007
predict	Regression results for new data	09-06-2008
	based on RPCR or RSIMPLS analysis	
Classical multivariate analysis and regression.		
ols	Ordinary (multiple) linear least squares regression	22-04-2004
mlr	Multivariate (multiple) linear regression	22-04-2004
classSVD	Singular value decomposition if more cases than variables	30-06-2003
kernelEVD	Singular value decomposition if less cases than variables	30-06-2003
cda	Classical linear and quadratic discriminant analysis	22-04-2004
cpca	Classical principal component analysis	30-06-2003
cpcr	Classical principal component regression	30-06-2003
csimca	Classical soft independent modelling of class analogies	20-09-2004
csimpls	Partial least squares regression (SIMPLS)	30-06-2003
Clustering methods.		
agnes	Agglomerative Nesting	20-10-2006
clara	Clustering method for Large Applications	20-10-2006
clusplot	Bivariate clustering plot of output from pam, fanny or clara	20-10-2006
daisy	Computing pairwise dissimilarities	20-10-2006
diana	Divisive Analysis	20-10-2006
fanny	Fuzzy Analysis	20-10-2006
mona	Monothetic Analysis	20-10-2006
pam	Partitioning Around Medoids	20-10-2006
tree	Tree plot for the output of agnes or diana	20-10-2006

Name	Description	Available since
Plot functions.		
makeplot	PlotGUI which includes the following plot functions:	30-06-2003
chiqqplot	Quantile-Quantile-plot of a vector	22-04-2004
	against the square root of the χ^2-quantiles	
ddplot	Robust distances versus Mahalanobis distances	22-04-2004
distplot	Plots a vector of distances	22-04-2004
ellipsplot	Scatter plot of bivariate data with 97.5% tolerance ellipse	22-04-2004
lsscatter	Scatter plot of bivariate data with regression line	22-04-2004
normqqplot	Quantile-Quantile plot of a vector against	22-04-2004
	the quantiles of a standard normal distribution	
daplot	Scatter plot of grouped bivariate data with their	22-04-2004
	97.5% tolerances ellipses (estimated from a discr. analysis)	
regresdiagplot	Regression diagnostic plot	30-06-2003
	(residual distance versus score distance)	
regresdiagplot3D	3D diagnostic plot	30-06-2003
	(residual distance versus score distance and orth. distance)	
residualplot	Plots the residuals from a regression analysis	22-04-2004
screeplot	Plots eigenvalues or their logarithm	30-06-2003
scorediagplot	Score diagnostic plot	30-06-2003
	(orthogonal distance versus score distance)	
simcaplot	Scatter plot with boundaries defined by	20-09-2004
	the number of principal components (estimated from simca)	

Name	Description	Available since
Functions used as subroutines and which can make life easy.		
greatsort	Sorts a vector in descending order	30-06-2003
heatmap	Displays a matrix as a heatmap image (downloaded from Mathworks.com)	28-06-2016
mahalanobis	Computes the distance of an observation with respect to the location and the shape of the data	22-04-2004
mcenter	Mean-centers a data matrix	30-06-2003
plotnumbers	Puts index of observations on a plot	30-06-2003
putlabel	Puts labels of observations on a plot	30-06-2003
randomset	Randomly draws a subset	09-06-2008
removal	Deletes rows/columns from a matrix	30-06-2003
robstd	Columnwise robust standardization	22-04-2004
twopoints	Generates directions through two data points	09-06-2008
uniran	Random uniform generator	30-06-2003
weightmecov	Weighted mean and covariance matrix	17-12-2004
Functions used only as subroutines.		
cvMcd	Cross-validated PRESS value for the MCD method	20-09-2004
cvRobpca	Cross-validated PRESS value for the ROBPCA method	20-09-2004
cvRpcr	Cross-validated RMSE value for the RPCR method	17-12-2004
cvRsimpls	Cross-validated RMSE value for the RSIMPLS method	17-12-2004
extractmcdregress	Auxiliary function for cross-valid. with RPCR and RSIMPLS	17-12-2004
removeObsMcd	Removal of observations for calculation of PRESS (used in cvMcd)	20-09-2004
removeObsRobpca	Removal of observations for calculation of PRESS (used in cvRobpca, cvRpcr, cvRsimpls)	20-09-2004
robpcaregres	Robust regression based on results from ROBPCA (used in rsimpls and cvRsimpls)	17-12-2004
rrmse	Robust RMSECV and RMSEP values (used in rpcr and rsimpls)	30-06-2003
rsquared	Robust and classical R^2 values	30-06-2003
rstep	Reflection step (used in rapca)	30-06-2003

Datasets

Datasets from the book *Finding groups in data: An introduction to cluster analysis*, Kaufman L. and Rousseeuw P.J., Wiley, New York, 1990:
agricul.mat, animal.mat, country.mat, flower.mat, obj200.mat, ruspini.mat.

History and major updates

Release June 30, 2003

The toolbox is made available with main functions: mcdcov, rapca, robpca, ltsregres, mcdregres, rpcr, rsimpls.

Release April 22, 2004

Several robust and classical procedures have been added:

- robust estimators of location and scale (M-estimators, Hodges-Lehmann, ...)
- the medcouple: a robust estimator of skewness
- robstd: robust standardization of multivariate data
- rda/cda: robust and classical discriminant analysis (classification)
- ols, mlr: classical least squares regression

Moreover several of the main functions are updated:

- mcdcov,rapca,ltsregres: the input and output structure is made conform to that of robpca, rpcr,...
- ltsregres: the intercept adjustment is now made optional. In the default setting, no adjustment is performed to save computation time. Also in mcdcov, some improvements have been made to speed up the computations.

Release September 20, 2004

Several robust and classical procedures have been added:

- csimca/rsimca: classical and robust SIMCA
- pressmcd/pressrobpca/removeobsmcd/removeobsrobpca/updatecov: subroutines to use in fast cross-validation methods for MCD en ROBPCA.

Updates of some of the main functions were performed:

- makeplot: accompanying plots for csimca, rsimca, were added
  Classical plots will now automatically be plotted if classical output is provided.

Release December 17, 2004

- Cross-validation for robust calibration methods (RPCR, RSIMPLS) has been added. The 'pressmcd' and 'pressrobpca' auxiliary functions are renamed into 'cvMcd' and 'cvRobpca'. To select the appropriate number of latent variables, several graphical displays are added, among which the Robust Component Selection (RCS) curve.
- The classification functions (cda, rda, csimca, rsimca) allow an extra argument: a prediction set, different from the training set, on which the classification rules are applied.

Release March 23, 2005

LIBRA now also works with MATLAB version 7.0. Reported bugs have been fixed (especially in the function makeplot.m) and some minor updates were performed on the functions: robpca, rsimpls, rrmse, cvMcd.

Release October 20, 2006

LIBRA includes the clustering algorithms described in the book *Finding groups in data: An introduction to cluster analysis* of Kaufman and Rousseeuw (Wiley, 1990).

Release December 22, 2006

The function to compute and plot a skewness adjusted boxplot has been added.

Release March 05, 2007

- Corrected bug in *mcdcov*: correlation matrix of classical analysis.
- Corrected bug in *rpcr*: reporting of RCS values
- Updated the function *weightmecov* such that it is less memory exhaustive.

Release July 31, 2007

The function to compute censored depth quantiles has been added.

Release February 28, 2008

- Added the functions: *adjustedoutlyingness.m* and *bagplot.m*.
- Corrected bug in *fanny*: lines 135-140 added initialisation of the vector 'dv'.

Release March 27, 2008

The function ROBPCA has an additional input argument *skew* which allows to perform robust PCA for skewed data.

Release April 21, 2008

Bugs corrected in

- cvrsimpls (line 328: resrob.flag.all)
- robpca (line 561 + 628: kmax back in the output)
- rsimpls (line 347: out.weights2=out.robpca.flag.all)

Release June 9, 2008

- Added a new function *predict*: computes regression results for new data based on the output from a RPCR or RSIMPLS analysis.
- robpca, rapca, cpca: out.flags extended to

```
out.classic.flag.od=(out.classic.od<=out.classic.cutoff.od);
out.classic.flag.sd=(out.classic.sd<=out.classic.cutoff.sd);
out.classic.flag.all=(out.classic.flag.od)&(out.classic.flag.sd);
```

- rsimpls, rpcr, csimpls, cpcr: out.flags extended to

```
out.flag.od=out.od<=out.cutoff.od;
out.flag.resd=abs(out.resd)<=out.cutoff.resd;
out.flag.all=(out.flag.od & out.flag.resd);
```

- rsimpls: extra output argument introduced: the covariance matrix of the scores T, out.Tcov, needed for the *predict* function.
- Adjustedoutlyingness: bug corrected, and separate functions created: *twopoints* and *randomset*

Release June 12, 2009

- The figures for the cluster programs can now also be obtained via the makeplot function.
- Mex-files included (instead of the older .dll) functions to call compiled C-code.

Release August 27, 2009

In *pam*, the average silhouette width per cluster is now correctly computed.

Release November 06, 2009

In *ltsregres* a small bug is corrected for small data sets with ties.

Release October 30, 2011

- New functions added: *classifskew* and *DetMCD*
- Small bugs corrected in *ltsregres*, *kernelEVD*, *classSVD*, *adjustedboxplot*, *unimcd*

Release June 28, 2016

- New functions added: *ddc*, *cellmap* and *heatmap*
- Several functions have been extended to allow a fixed zero center: *mcdcov*, *robpca*, *weightmecov*, *classSVD*, *kernelEVD*
- Small bugs corrected in *agnes*, *DetMCD*, *twopoints*