

Metabolomics profiling statistical analysis (MOPSA) V 1.0

User manual

1. Introduction

Metabonomics is an emerging field providing new insights into physiological processes and difference. Besides conventional PCA, PLS and OPLS approaches, more and more machine learning classifiers are likely to become the supplements for metabolic profiling data analysis.

A comprehensive comparison of PLS, support vector machine (SVM, with linear and quadratic kernels), linear discriminant analysis (LDA), and random forest (RF) was conducted applying on clinical metabonomics data. The overall performance of RF and SVM (linear kernel) is superior to the others.

The Matlab code for Partial least squares analysis (PLS), Support vector machines (SVM), Linear discriminant analysis (LDA), and Random forest (RF) and for performance comparison (Permutation, k-fold/hold-out cross validation, and ROC) are integrated and a GUI was developed. The software is protected by CPCC and freely available for academic usage.

2. How to use

1) Startup

Open Matlab 7.1, go to the directory of MOPSA and run the file “MOPSA.m” (fig. 1). The main window of MOPSA will appear (Fig. 2).

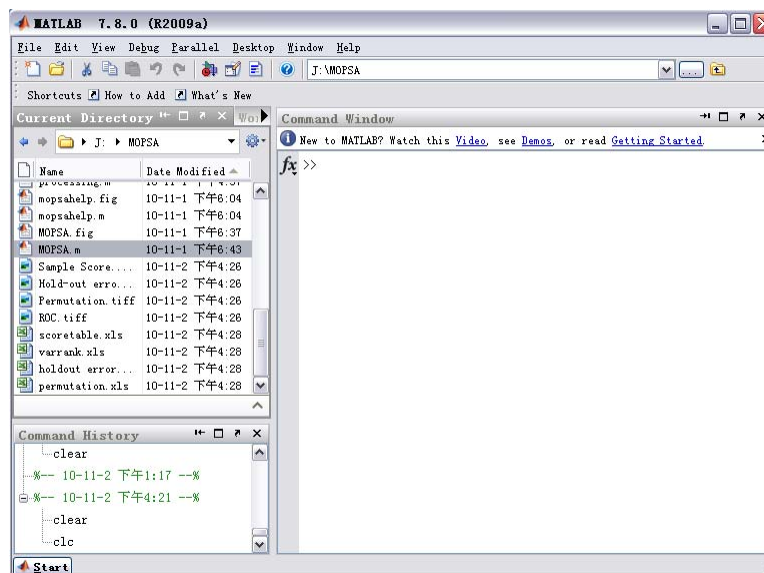


Fig. 1 Startup

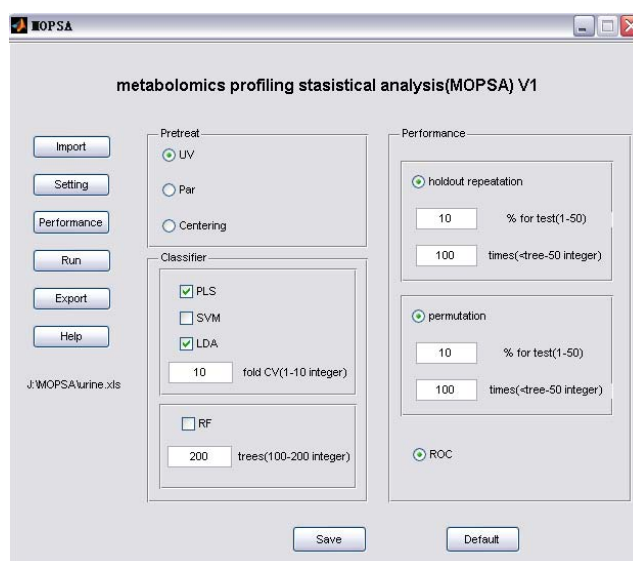
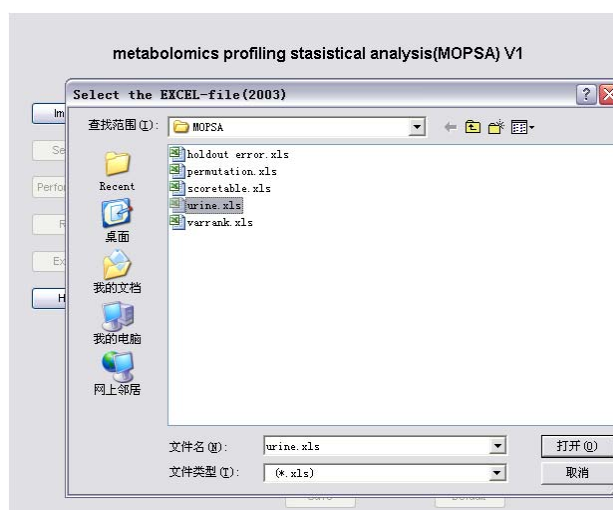


Fig. 2 The main window of MOPSA

2) Data importation

Click “Import” button and select the .xls file to import (Fig. 3-4).



	A	B	C		A	B	C	D	E	F	G
1	SN	样本编号 (整数)			1		2	3	4	5	6
2	Y	样本类别 (1或者2)				1	1	1	2		2
3	变量编号 (整数)	变量强度	变量强度		1	2.969	3.4384	5.5893	1.4101	2.0138	3.0222
4		变量强度	变量强度		2	41.673	225.46	40.739	43.561	55.92	7.832
5		变量强度	变量强度		3	29.781	16.688	17.783	19.183	39.191	10.141

Fig. 3 Data file format for importation

3) Function selection

Select the data preprocessing (UV, Par, or Centering), analysis methods (PLS, SVM, and LDA), and performance (cross validation, permutation, and ROC) manner. Set the parameters or use the defaults.

4) Result

Click “run”, the main window will disappear and the software will run in the background (Fig. 4).

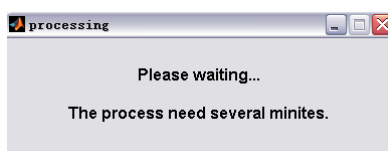


Fig. 4 Running

After running, representative plots will be saved in the same folder of dataset and will be shown automatically (Fig. 5).

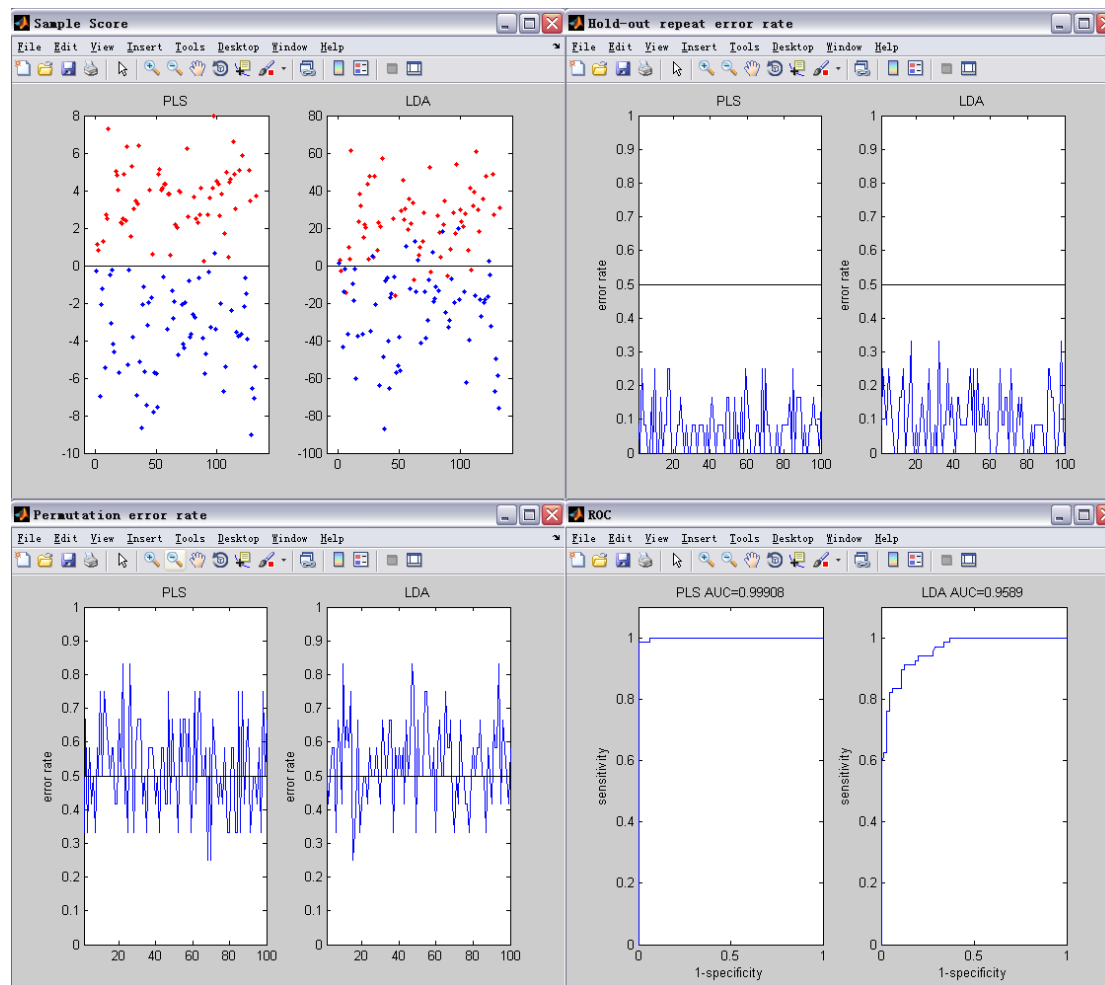


Fig. 5 Result example

Click the “export” button to save the tables: scoretable.xls (scores for classification), varrank.xls (VIP values provided by classifiers), holdout error.xls (holdout repeat errors), and permutation.xls (permutation errors).