RLS2 MATLAB Toolbox Tutorial

RLS2 MATLAB Toolbox is a set of scripts that implements RLS2 (regularized least squares with two layers) and RLS2LIN (linear regularized least squares with two layers). RLS2 is an instance of multiple kernel learning algorithm that can be used to simultaneously learn a regularized predictor and the kernel function in the form:

$$f(x) = b + \sum_{i=1}^{\ell} c_i \sum_{j=1}^{m} d_j \widetilde{K}_j(x_i, x),$$

where $\widetilde{K}_j$ are positive kernels. RLS2 is also an instance of kernel machine with two layers that extends the classic regularized least squares algorithm. RLS2LIN implements a version of RLS2 specialized to linear kernels on each feature:

$$\widetilde{K}_j(x, y) = x^j y^j.$$

RLS2LIN simultaneously performs regularization and linear feature selection, is memory efficient and very well suited for datasets with a large number of features. The package contains a Graphic User Interface (GUI) to load data, perform training, validation and plot results. The features of the GUI interface include:

- Data pre-processing.
- Efficient regularization path computation.
- Cross-validation.
- Random splits.
- Hold-out set validation.
- Approximate degrees of freedom computation.
- Plot results and export figures to PDF format.

Installation

To install the toolbox, simply unpack the package `RLS2.zip` into some folder and add that folder to the MATLAB path, by selecting "Set Path..." from the

File menu. To run the GUI, execute `rls2tools` from the MATLAB command line. To enable the export to PDF feature of the GUI, you need to download the script from save2pdf.m[1] and save it into the Matlab path.

Data format (command line)

- `X`: an $\ell \times n$ numerical matrix containing **input patterns**. Sparse arrays are also supported.
- `Y`: an $\ell \times m$ numerical matrix containing **output** values. For binary classification just label the classes with ± 1. For one versus all multi-class classification with m classes, use m sparse columns of binary indicators.

Execute `help rls2` or `help rls2lin` from the MATLAB command line for further information.

Data format for the GUI

The Graphic User Interface can be used to load data files in .MAT workspace format. The .MAT file must contain the following variables:

- `X`: an $\ell \times m$ numerical matrix containing **input patterns**. Sparse arrays are also supported.
- `Y`: an $\ell \times m$ numerical matrix containing **output** values **(regression)**.
- `Y`: an $\ell \times 1$ nominal array containing **output** labels **(classification)**. Execute `help nominal` from the MATLAB command line for further information.
- `T`**(optional)**: an $\ell \times 1$ logical vector whose entries are **true** in correspondence with validation examples and **false** in correspondence with training examples. Can be used to provide a pre-defined validation set.

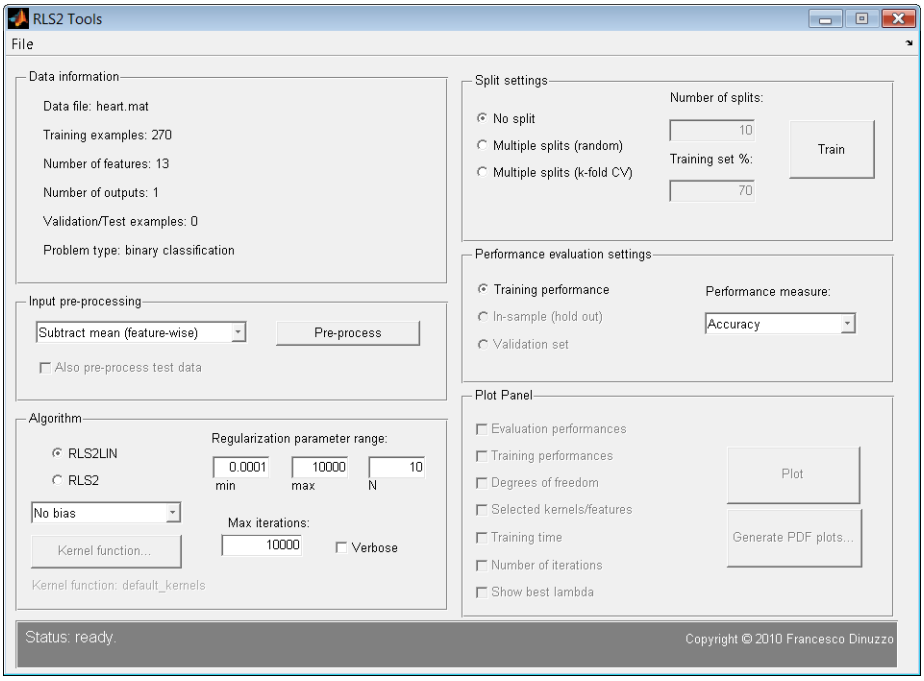
The GUI automatically determines the problem type (regression, classification) on the basis of the class of `Y`. The GUI also converts the data to command-line format upon loading.

[1] `http://www.mathworks.nl/matlabcentral/fileexchange/16179-save2pdf`

Tutorial 1

Follow this simple tutorial to quickly familiarize with the GUI and RLS2LIN:

- Make sure the toolbox is correctly installed.
- Execute `rls2tools` from the MATLAB command line to visualize the GUI.
- Choose the MATLAB current directory as the installation directory of RLS2 MATLAB toolbox.
- Select LOAD DATA... from the FILE menu.
- Choose the file `data/heart.mat` from the dialog box.

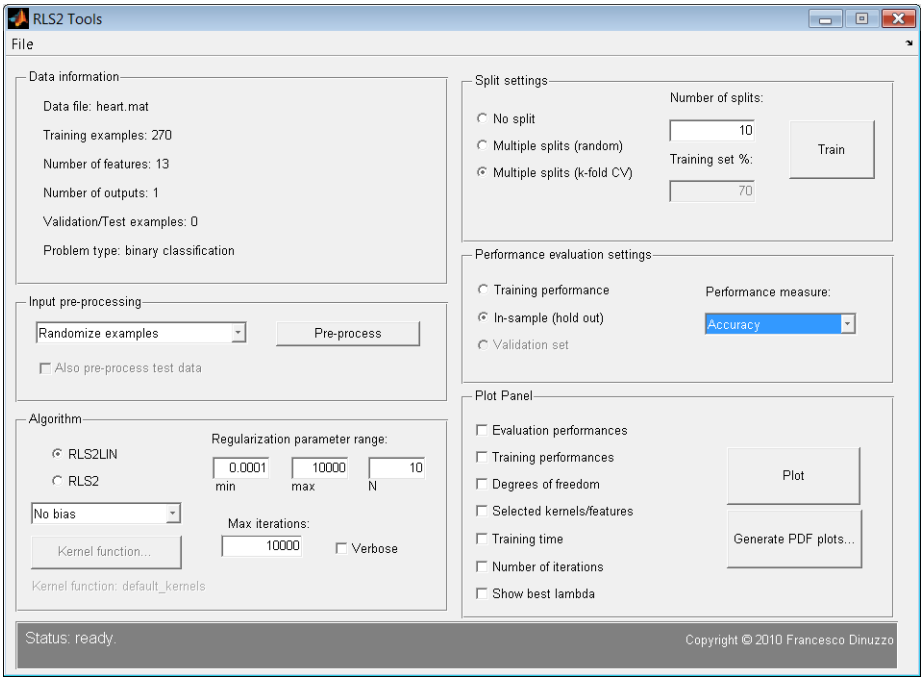


- If the dataset loads correctly, the DATA INFORMATION panel will contain basic information such as the file name, the number of examples, and the number of features. The GUI will automatically determine the problem type (in this case, classification) on the basis of the output data type (see Data format for the GUI).
- Before starting the training, you can perform some simple data pre-processing from the INPUT PRE-PROCESSING panel. Choose STANDARDIZE (FEATURE-WISE) from the popup menu, then click on the PRE-PROCESS button to standardize all the features. Next, choose RANDOMIZE from the popup

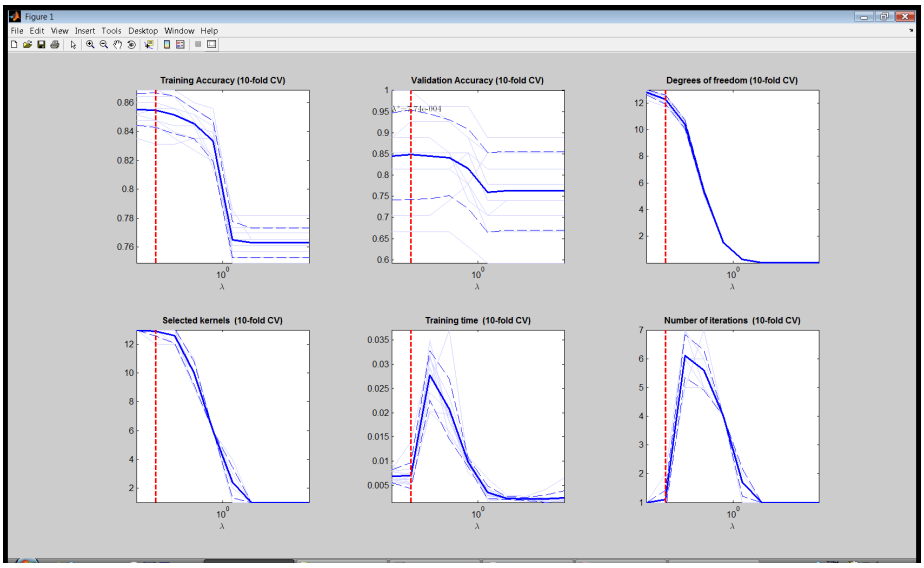
menu, and click again on the PRE-PROCESS button to randomly sort the examples.

- From the ALGORITHM panel, you can choose the learning algorithm (either RLS2 or RLS2LIN), a range (logarithmically spaced) for the regularization parameter and the maximum number of iterations. At this time, just use the default choices (RLS2LIN with default parameters).
- Click on the MULTIPLE SPLIT (K-FOLD CV) from the SPLIT SETTINGS panel. Use the default number of splits (10) to perform a 10-fold cross-validation.
- Click on the TRAIN button.

After the training, a new data structure called `models` will appear in the workspace, containing a set of RLS2 or RLS2LIN models. In addition, the PLOT PANEL will activate.



- Check all the boxes in the PLOT PANEL and click on the PLOT button to generate the plots. A new data structure called `statistics`, containing the plotted data, will appear in the workspace.



Tutorial 2

In this second tutorial, we will apply non-linear RLS2 to a simple regression problem with pre-defined validation set.

- Make sure the toolbox is correctly installed.
- Execute `rls2tools` from the MATLAB command line to visualize the GUI.
- Choose the MATLAB current directory as the installation directory of RLS2 MATLAB toolbox.
- Select LOAD DATA... from the FILE menu.
- Choose the file `data/prostate.mat` from the dialog box.

- The DATA INFORMATION panel shows that the dataset is provided with a pre-defined validation set of 30 examples.

- Check the box ALSO PRE-PROCESS TEST DATA in the INPUT PRE-PROCESSING panel.
- Choose STANDARDIZE (FEATURE-WISE) and click on the PRE-PROCESS button to standardize all the features. Then choose RANDOMIZE and click on the PRE-PROCESS button to randomly sort the examples.
- Choose RLS2 from the ALGORITHM panel.
- Set the minimum regularization parameter to 10^{-6} and maximum to 1.

At this point, it is possible to provide a customized .M file function computing a cell array of kernel matrices. The default kernel function computes Gaussian RBF and polynomial kernels with a variety of parameters.

- Observe that the VALIDATION SET option in the PERFORMANCE EVALUATION SETTINGS panel is already active.
- Click on the TRAIN button.
- Choose MSE from the PERFORMANCE MEASURE menu.
- Check all the boxes in the PLOT PANEL and click on the PLOT button to generate the plots.