IRootLab Tutorials

Classification with PCA-SVM

Julio Trevisan – juliotrevisan@gmail.com

26th/July/2013

  
This document is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](http://creativecommons.org/licenses/by-nc-sa/3.0/).

Loading the dataset 1

Pre-processing 2

Optimization of SVM parameters 5

Creation of required objects 5

Grid search 8

Visualization of results 10

Visualization of the optimization log 10

Classification confusion matrix for best parameters 12

# Introduction

Sdkfjhsdfjksdh fjksdh fsdjk fhsdhe SVM classifier has tuning parameters that need to be optimized before fitting data. This tutorial will take you through the steps of finding these parameters and then getting a confusion matrix for the optimally tuned classifier.

# Loading data, pre-processing

Please follow steps 1 to 13 in the “Classification with SVM” tutorial

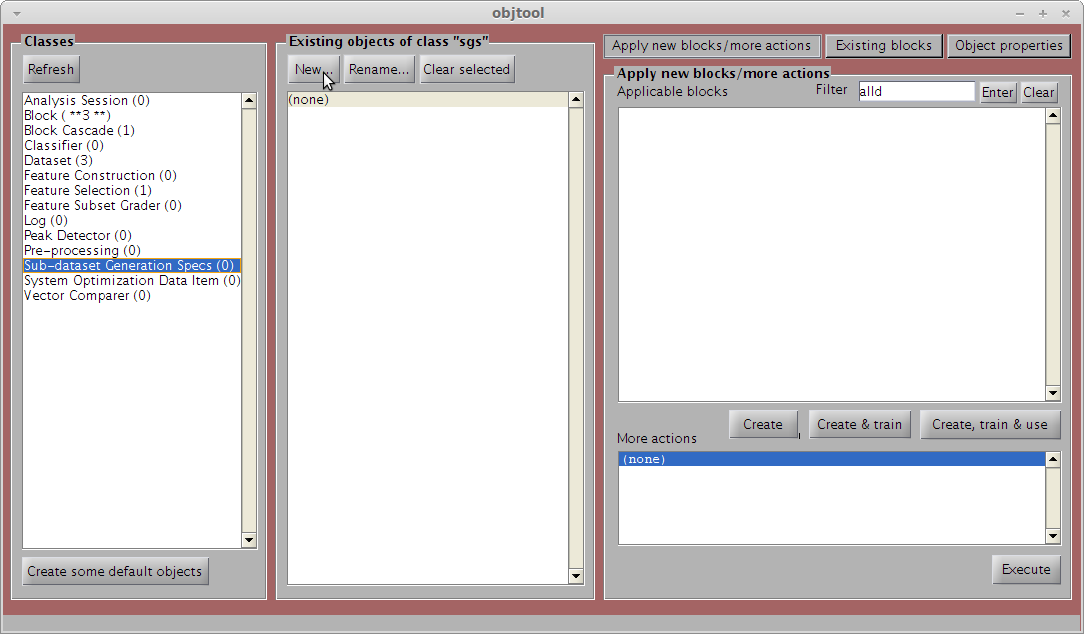
# Tuning the parameters

This tutorial utilizes the Gaussian kernel SVM, which implies that there are two parameters to tune: *c* and *gamma* (these parameters are referred to as *C* and *γ* in [1]). These parameters have to be tuned to the value that gives best classification. Apart from these two parameters, the number of principal components (PCs) has to be tuned as well. The number of PCs is added as an extra parameter in the grid search, so that the three parameters are tuned together.

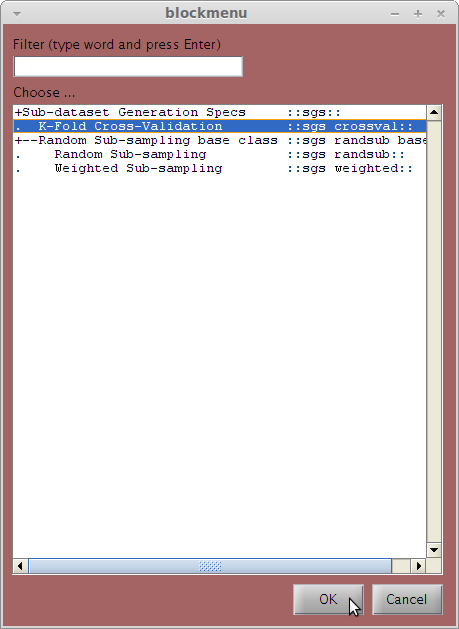
The optimization will use 5-fold cross-validation[2] to calculate the classification rates. The optimization technique is “grid search” as recommended[1].

## Creation of a “Sub-dataset generation specs” object

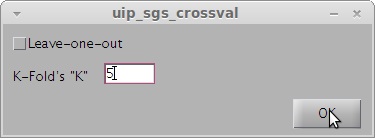
1. Click on “Sub-dataset Generation Specs” in left panel
2. Click on “New…” in middle panel



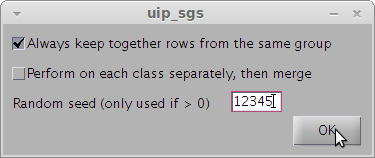
1. Locate and double-click on “K-fold Cross-Validation”



1. Enter “5” in the “K-Fold’s ‘K’” box

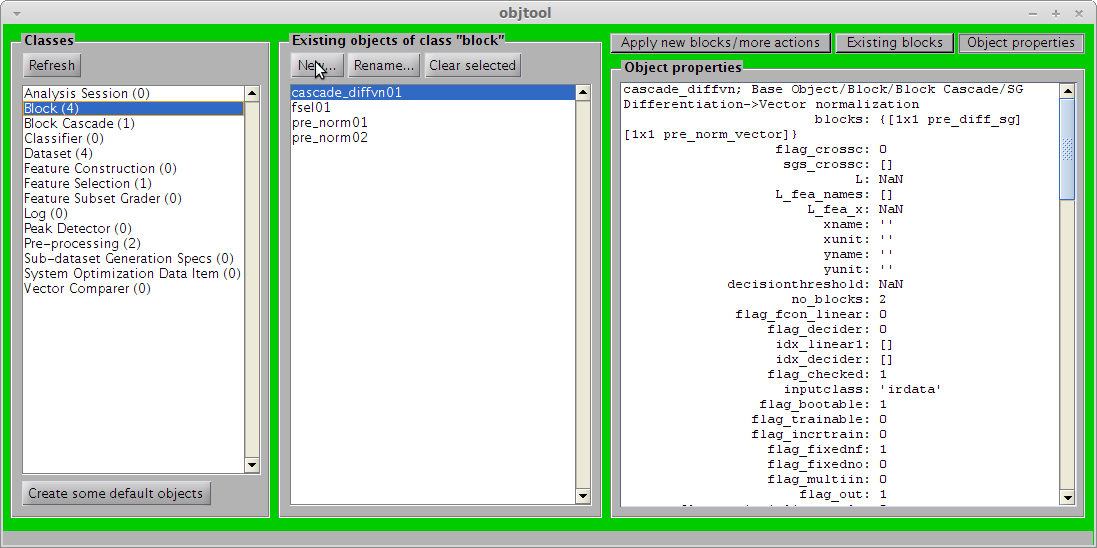


1. Optionally type any number (*e.g.*, 12345) in the “Random seed” box (recommended)
2. Click on “OK”

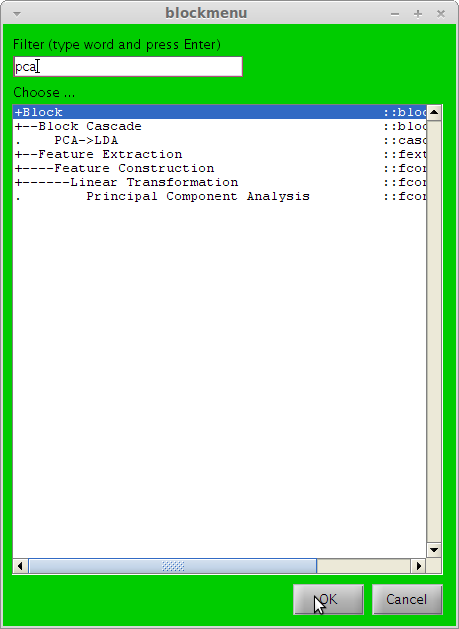


## Creation of the PCA-SVM block

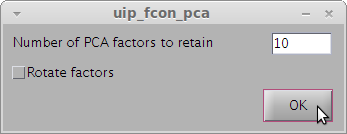
1. Click on “Block” in the left panel
2. Click on “New” in the middle panel



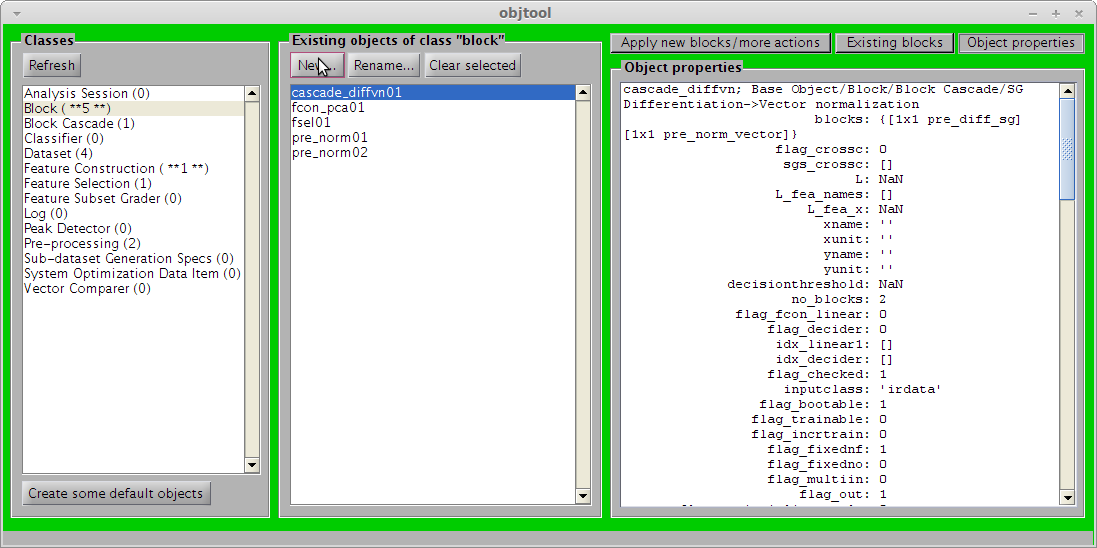
Locate and double-click on “Principal Component Analysis”



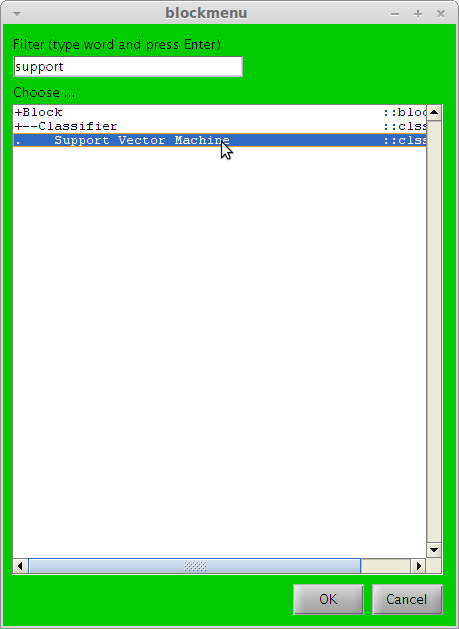
1. Click on “OK” (the number in the box is irrelevant at this point)



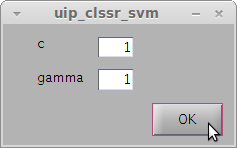
1. Click on “New…” again



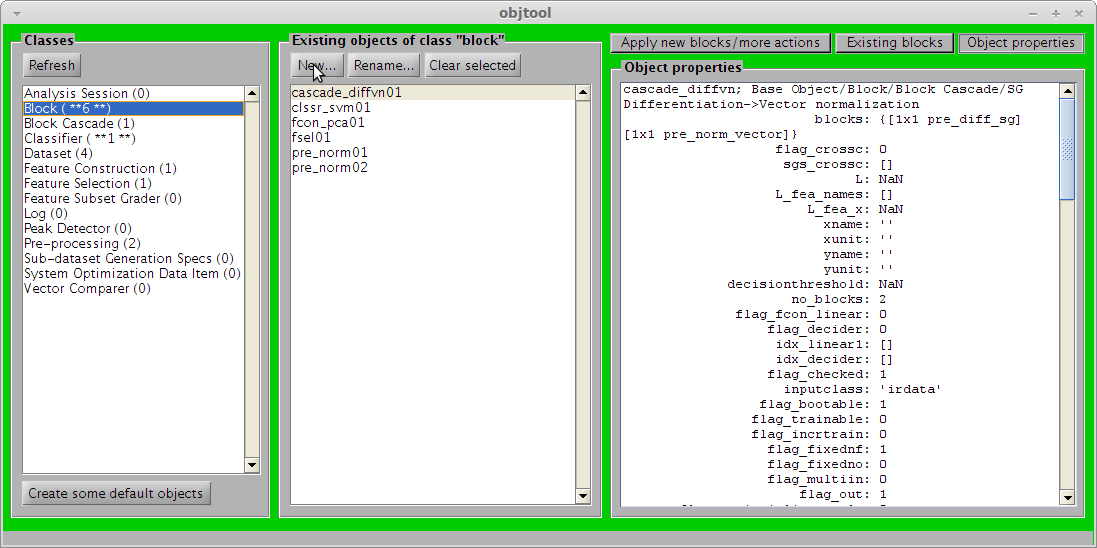
1. Locate and double-click on “Support Vector Machine”



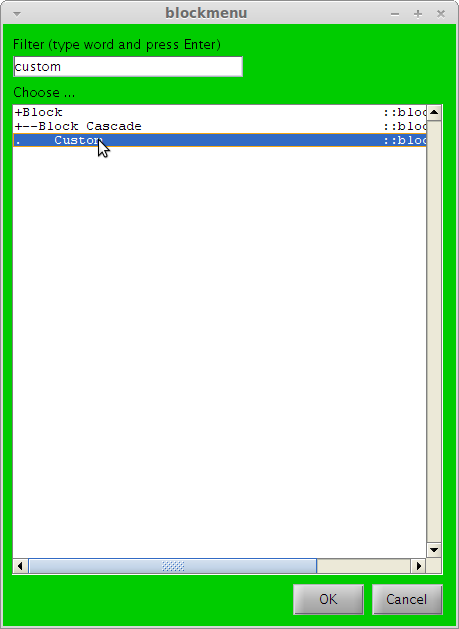
1. Click “OK” (the values in the boxes are irrelevant at this point)



1. Click on “New…” in the left panel once more

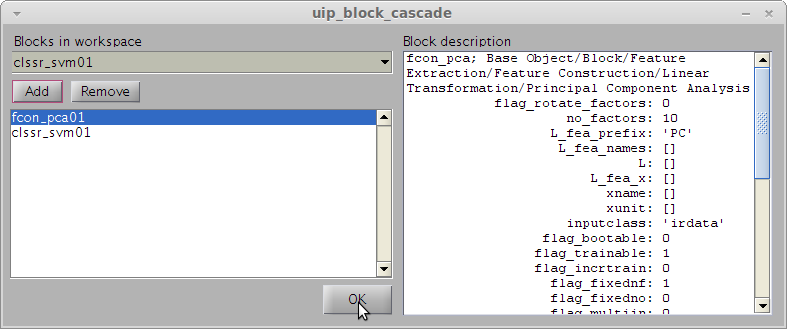


1. Locate and double-click on “Custom”



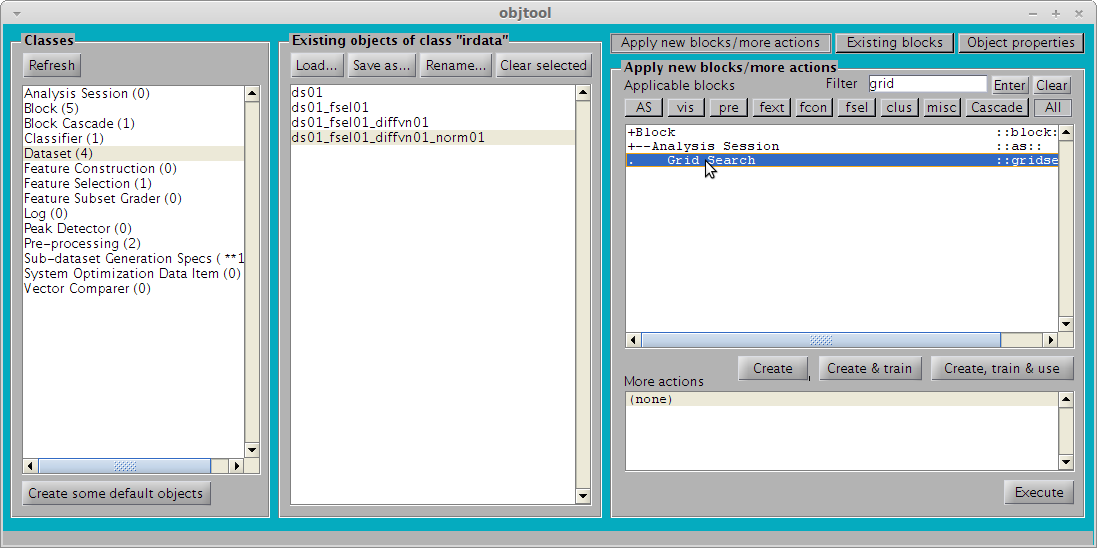
Now find and add the blocks named “fcon\_pca01”, and “clssr\_svm01” in this order

Click on “OK”

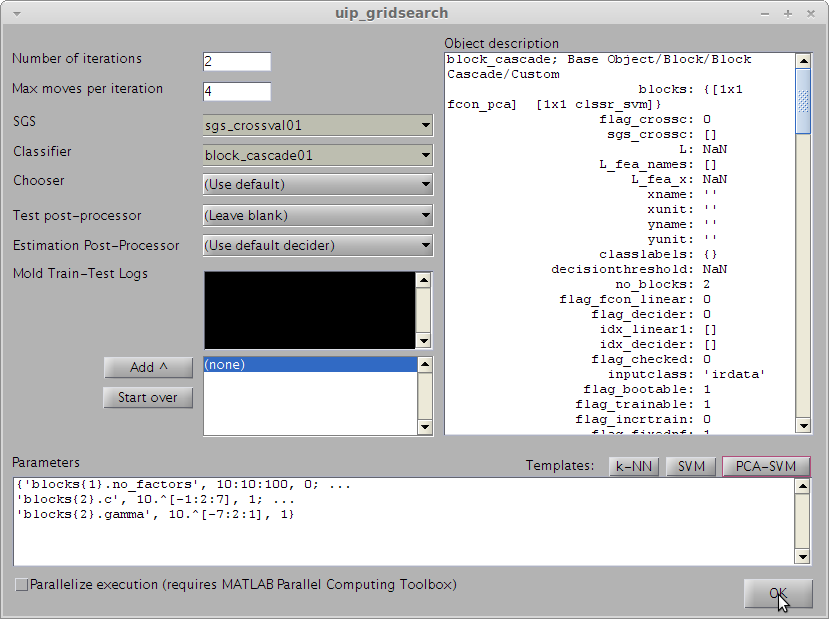


## Grid search

1. Click on “Dataset” in left panel
2. Click on dataset named “ds01\_fsel01\_diffvn01\_norm01” in middle panel
3. Locate and double-click “Grid Search” in right panel



1. In the “SGS” drop-down box, select “sgs\_crossval01”.
2. In the “Classifier” drop-down box, select “block\_cascade01”.
3. Click on the “PCA-SVM” button to fill in the “parameters” box appropriately.
4. You may optionally change the search ranges of the parameters.
5. Also, you may optionally check the “Parallelize execution” checkbox
6. Click on “OK”. **Warning:** this operation will be probably quite time-consuming (for example, for the settings in this tutorial it took around 3457s for the SHE dataset)

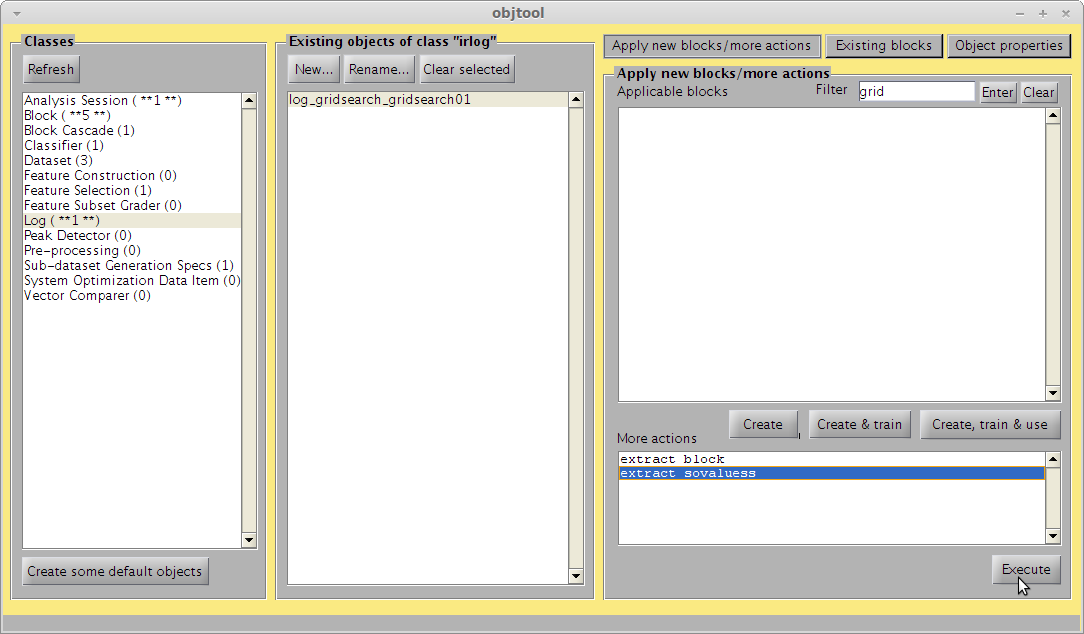


1. Watch MATLAB command window for progress indicator

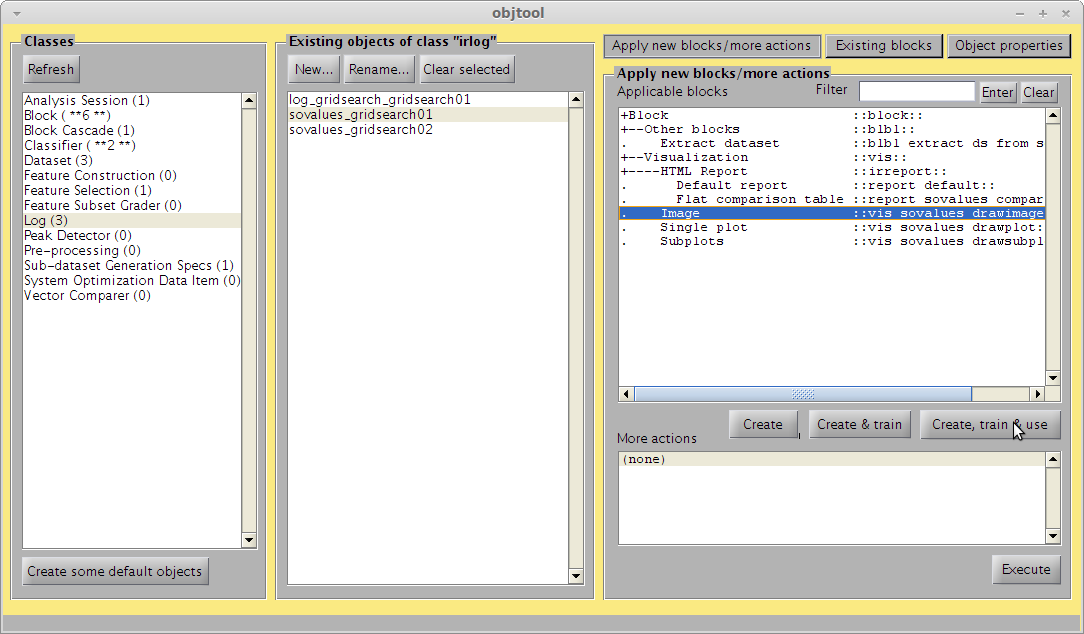
# Visualization of results

## Visualization of the optimization log

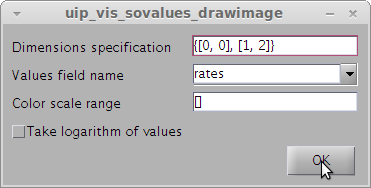
1. Click on “Log” in left panel
2. Select “log\_gridsearch\_gridsearch01” in middle panel
3. Double-click on “extract\_sovalues” in right panel

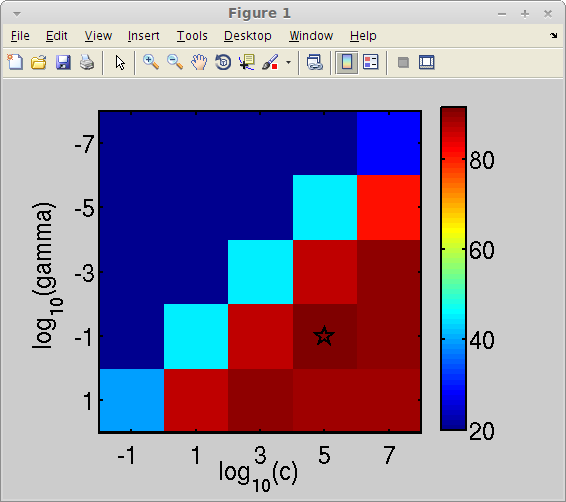


1. Click on “sovalues\_gridsearch01” in the middle panel
2. Locate and double-click on “Image” in the right panel

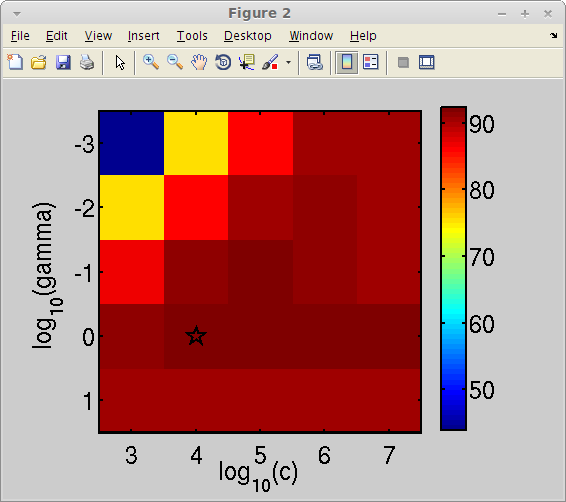


1. In the “Dimensions specification” box, change to “{[0, 0], [1, 2]}”
2. Click on “OK”



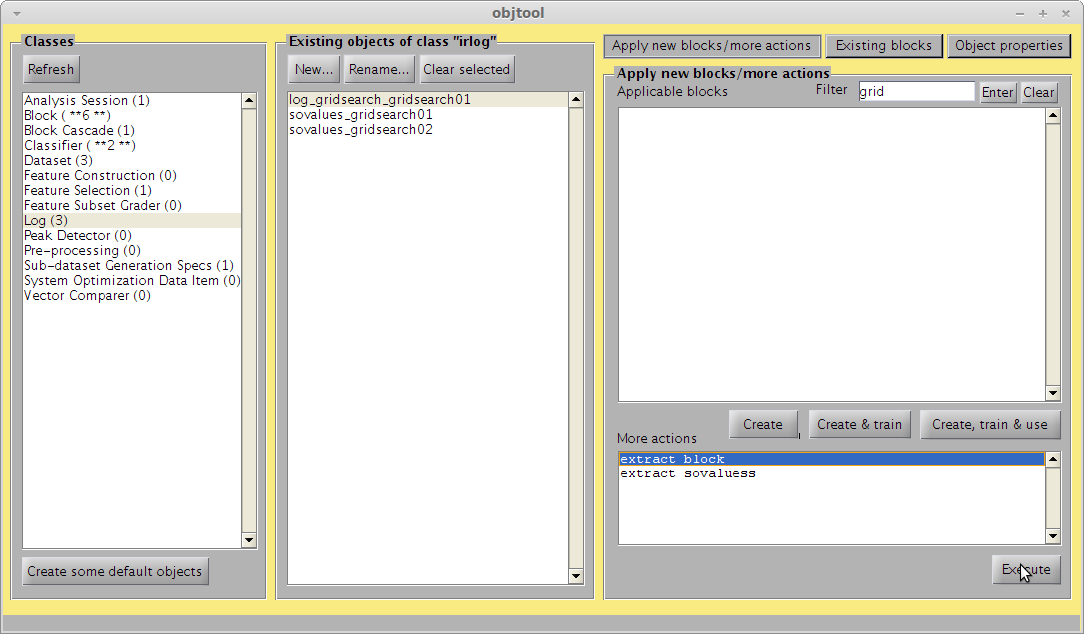


1. Repeat last 4 steps with the “sovalues\_gridsearch02” object in the middle panel

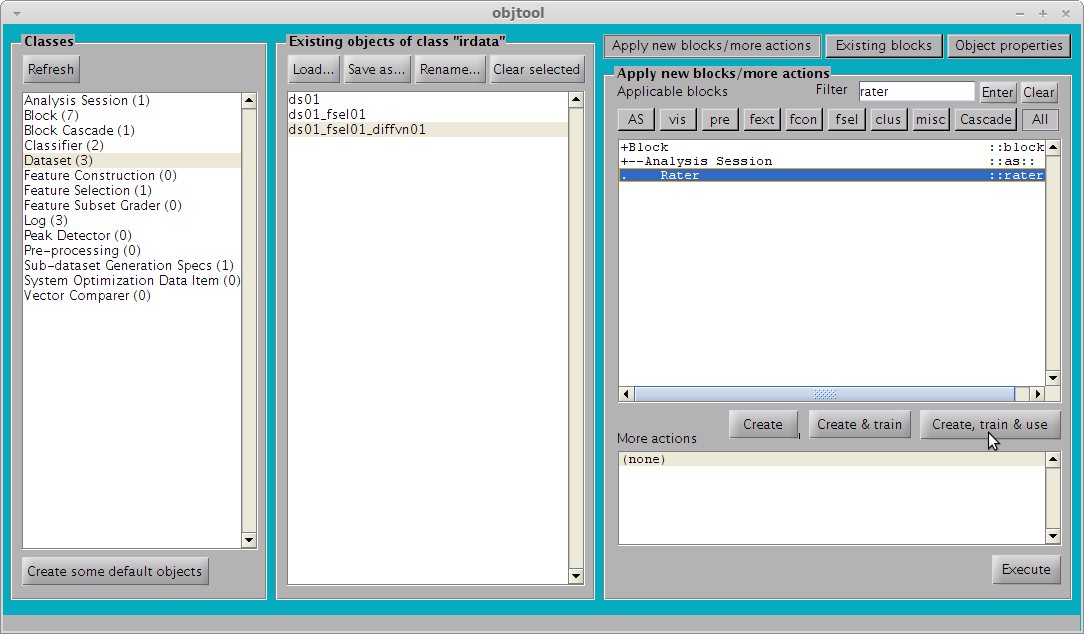


## Classification confusion matrix for best parameters

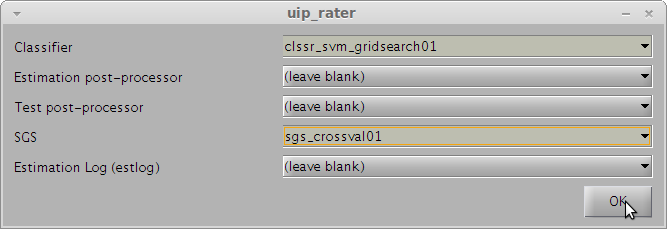
1. Click on “Log” in the left panel
2. Click on “log\_gridsearch\_gridsearch01” in the middle panel
3. Double-click on “extract\_block” in the right panel



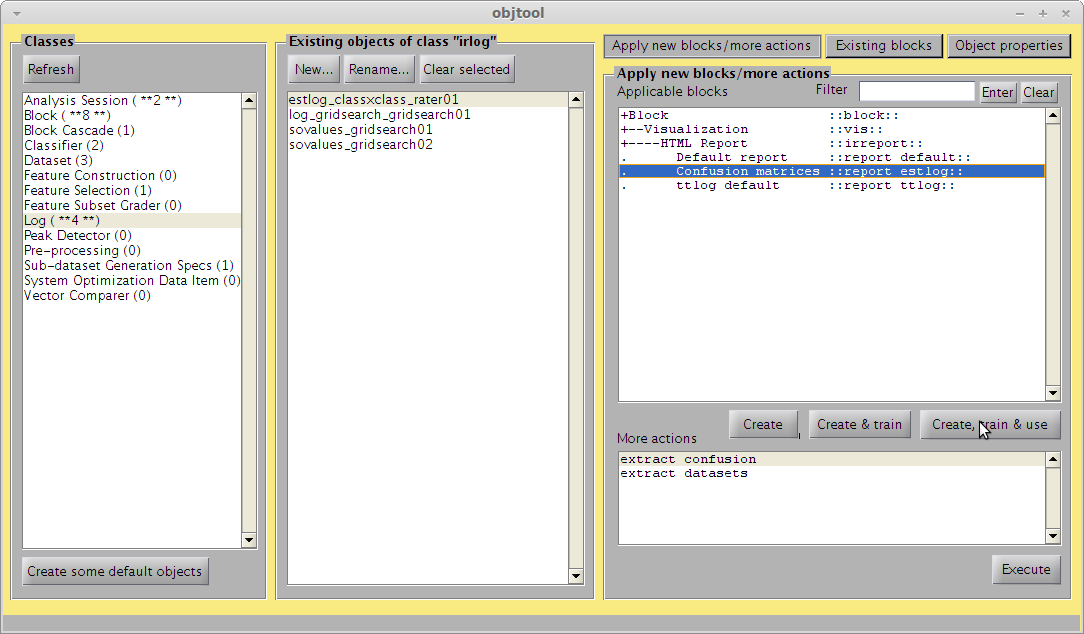
1. Click on “Dataset” in the left panel
2. Click on “ds01\_fsel01\_diffvn01\_norm01” in the middle panel
3. Double-click on “Rater” in the right panel



1. In the “Classifier” box, select “clssr\_svm\_gridsearch01” (this is the block that was created from the block extraction action above).
2. In the SGS box, select “sgs\_crossval01”. This will cause the cross-validated estimation to use the same dataset splits as the grid search optimization before.
3. Click on “OK”



1. Click on “Log” in the left panel
2. Click on “estlog\_classxclass\_rater01” in the middle panel
3. Double-click on “Confusion matrices” in the right panel



1. Click on “OK”

