Experiment Report

Name: RRLS_bench_icubdyn_RF_retuning
Number: 6
Date: 2014-07-17
Author: Raffaello Camoriano (raffaello.camoriano@iit.it)

Goal

Answer the following question:

- On the first 40000 samples of the iCubDyn dataset (approx. 15 minutes recording of random arm motion), does the periodic retuning of the lambda parameter every 10000 samples (3 min, 45 s) significantly affect the accuracy of the predictions?

Data

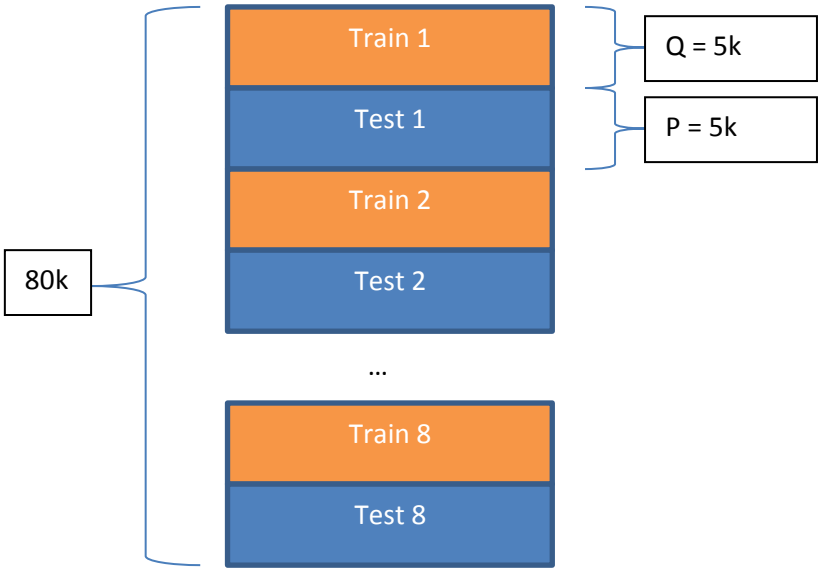
This experiment uses the iCubDyn dataset, projected over a 200-dimensional random features space. The RF projection used here is the standard GURLS one:

```
W = sqrt(2)*randn(d,numRF);
G = [cos(W*X);
     sin(W*X)];
```

The samples are not randomized, since batch solution is exact and the RF approximation is assumed to be sufficiently accurate, at least to a first approximation.

The dataset is split into 16 chunks, alternated between training and test. Each chunk is composed of 5000 samples. Each set is in the form of an input data matrix and an output labels vector.

```
q = 5000;    % Training subsets size
p = 5000;    % Test subsets size
```



Algorithms

1) Batch, no λ retuning

The employed batch algorithm is the standard one, documented in the GURLS manual (example in section 2.3.2).

1) Hyperparameter λ is selected among 1000 guesses via hold-out on *Train1*. Validation size: 20% of the training set. This λ is kept fixed throughout the experiment.
2) Training is performed from scratch on *Train_i* for i = 1:8, using the lambda selected in the beginning (step 1).
3) Testing is performed on *Test_i* for i = 1:8 and the results (RMSE and predictions) are stored.

2) Batch with λ retuning

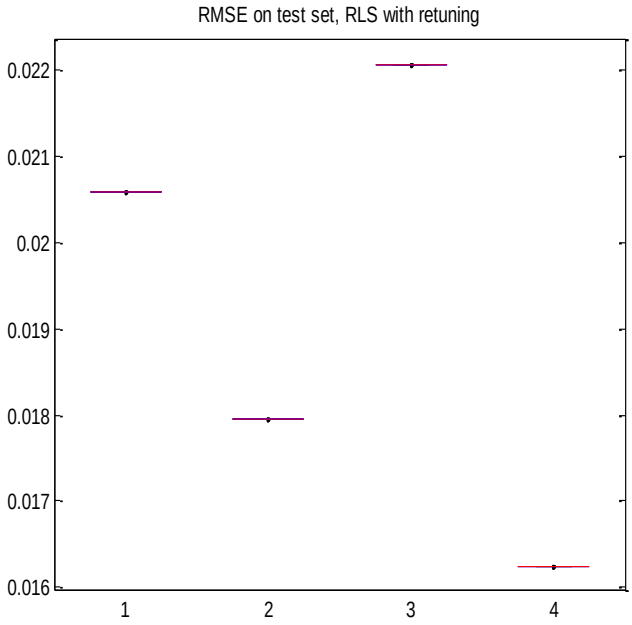
The employed batch algorithm is the standard one, documented in the GURLS manual (example in section 2.3.2).

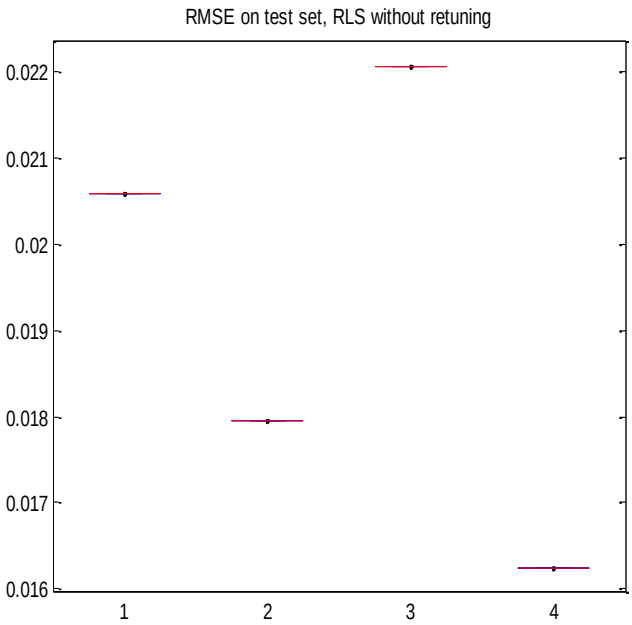
4) Hyperparameter λ is selected among 1000 guesses via hold-out on *Train_i*, for i = 1:8. Validation size: 20% of the training set. This λ is retuned for each training chunk.
5) Training is performed from scratch on *Train_i* for i = 1:8, using the most recent optimal lambda selected at step i.
6) Testing is performed on *Test_i* for i = 1:8 and the results (RMSE and predictions) are stored.

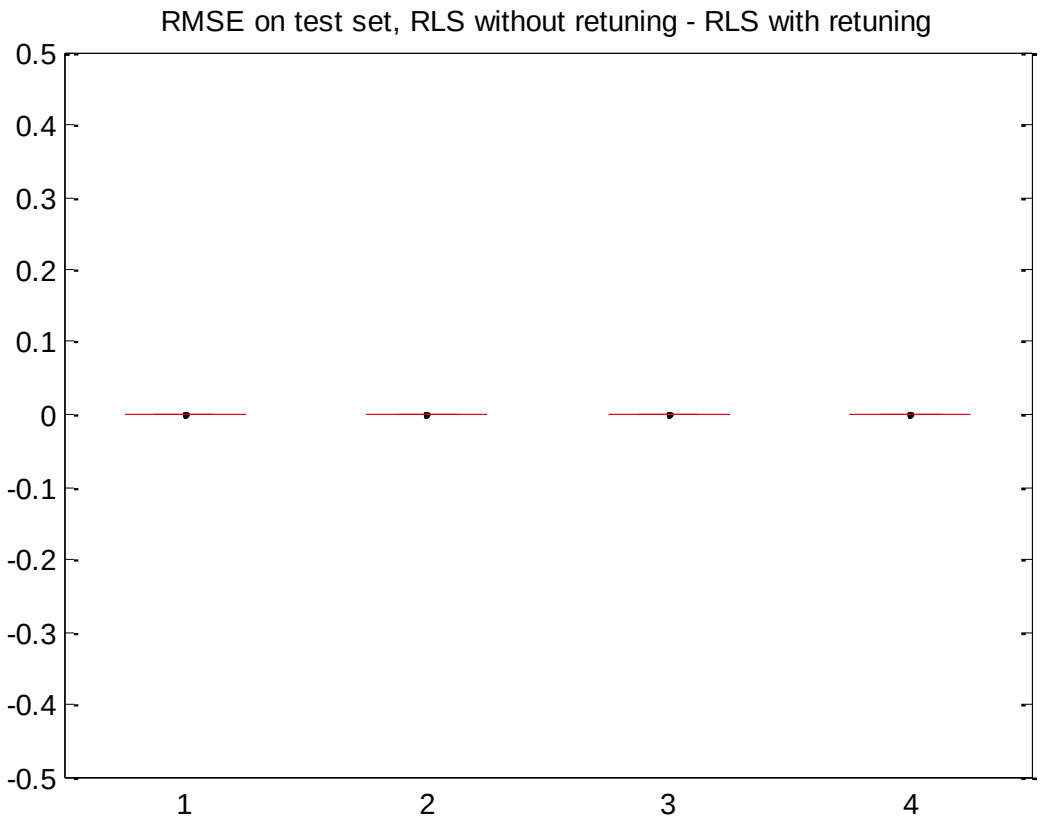
Results

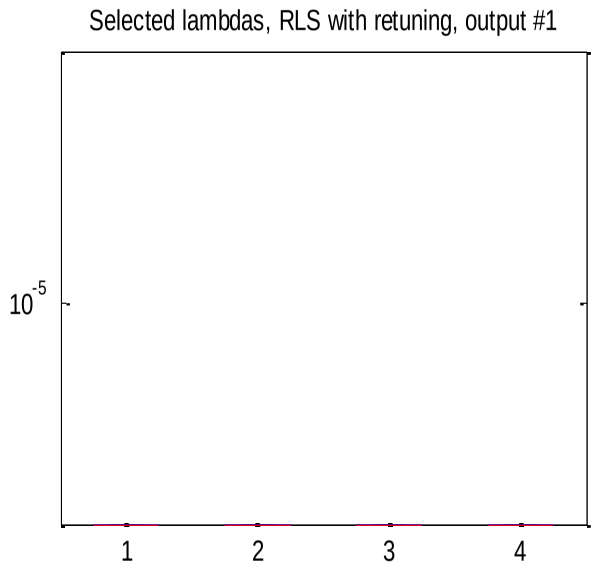
- The results reported here are referred to 20 independent runs of the experiment.
- The data has been normalized and scaled by factors 1, 10, 100, 1000. Scaling has been introduced to evaluate the possible effects of varying sigma on the choice of the optimal lambda and on the resulting RMSE.

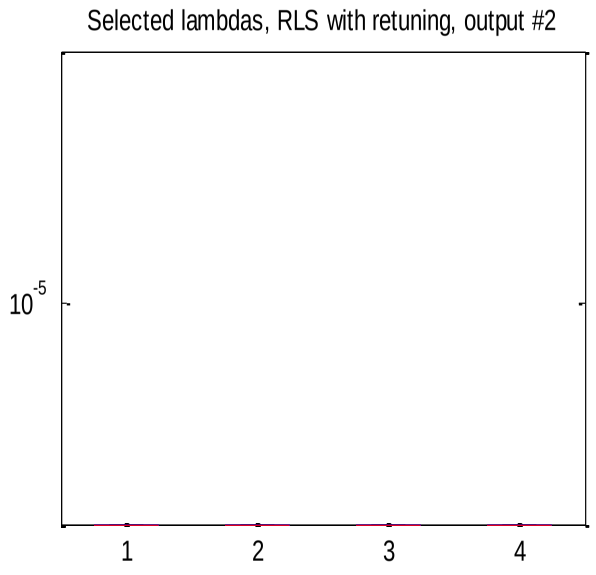
Scaling 1

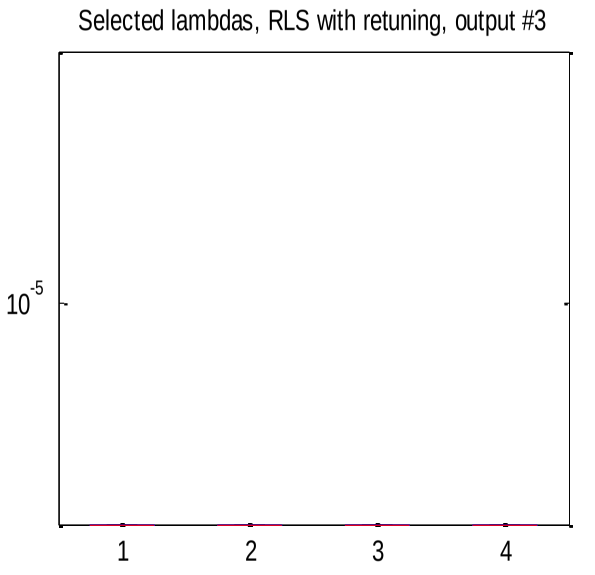

RMSE on test set, RLS with retuning
0.022
0.021
0.02
0.019
0.018
0.017
0.016
1
2
3
4

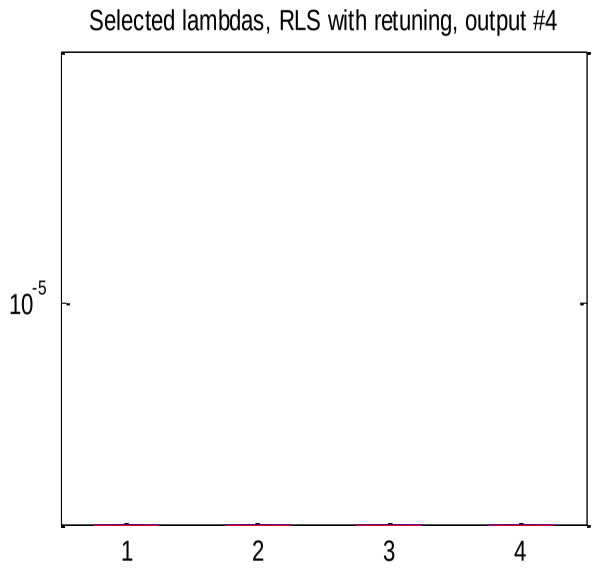

RMSE on test set, RLS without retuning
0.022
0.021
0.02
0.019
0.018
0.017
0.016
1
2
3
4

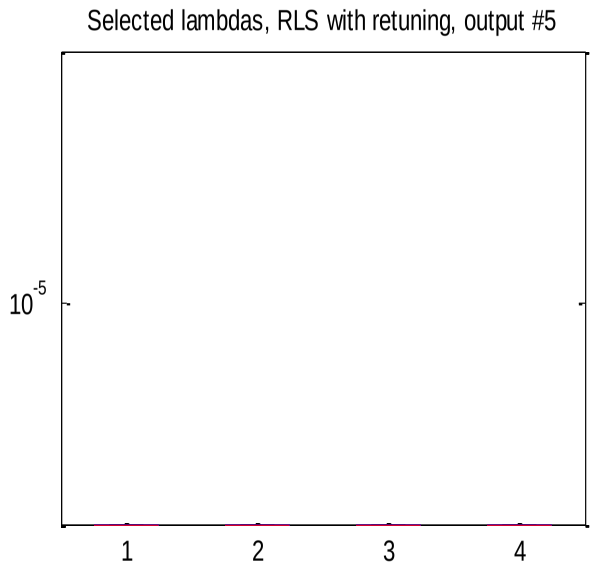

RMSE on test set, RLS without retuning - RLS with retuning
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
1
2
3
4

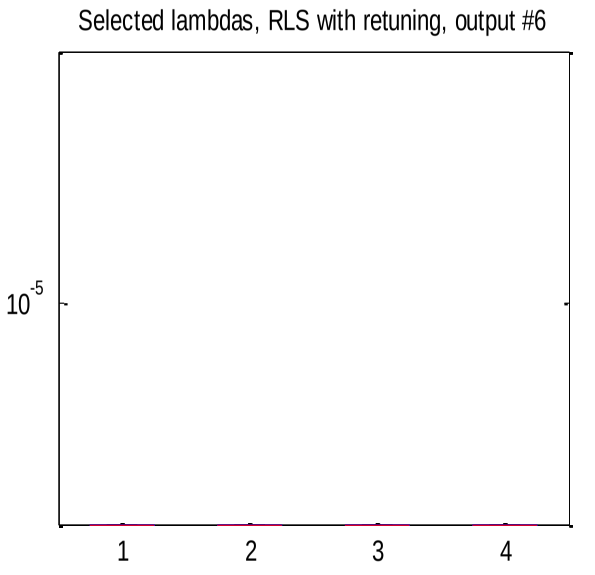

Selected lambdas, RLS with retuning, output #1
10^-5
1
2
3
4

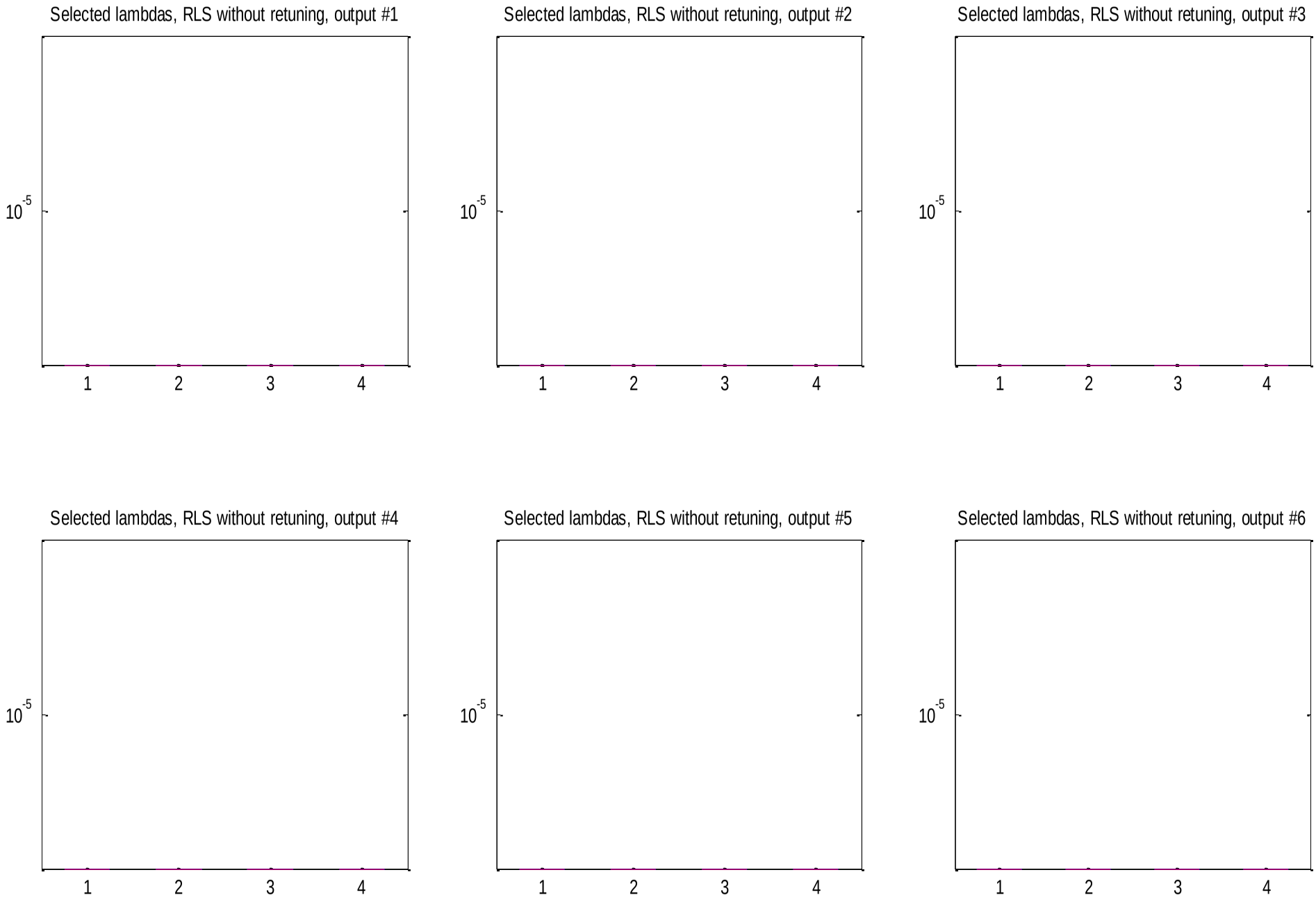

Selected lambdas, RLS with retuning, output #2
10^-5
1
2
3
4


Selected lambdas, RLS with retuning, output #3
10^-5
1
2
3
4

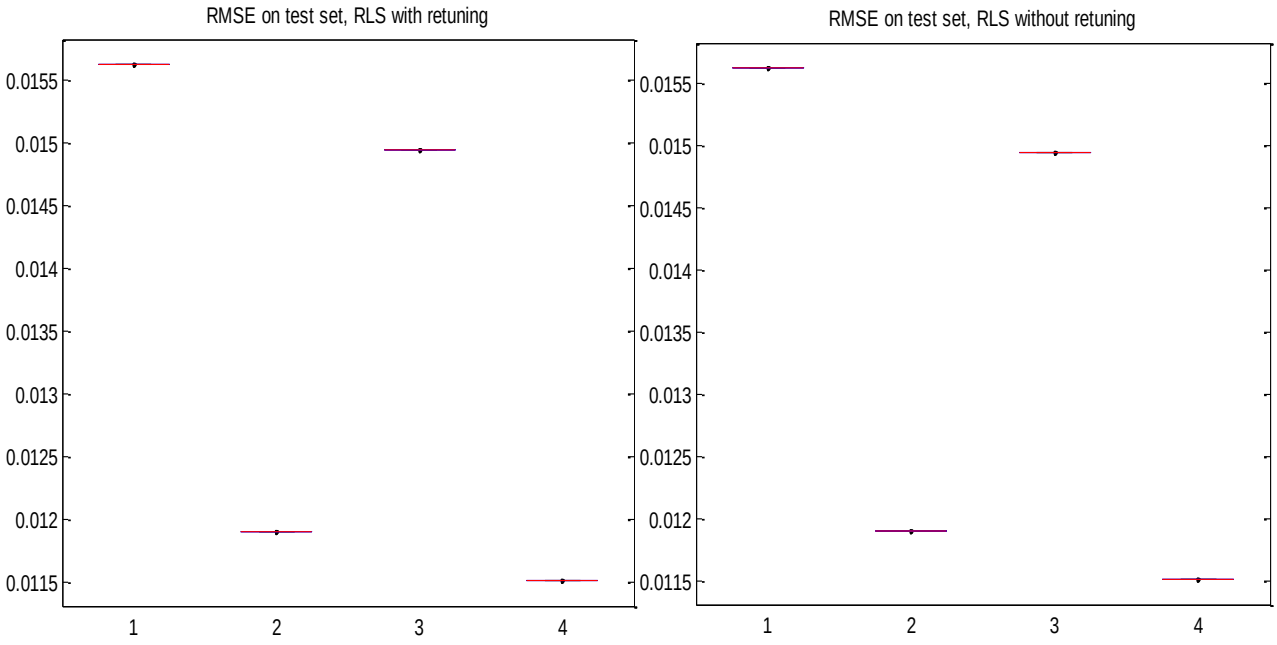

Selected lambdas, RLS with retuning, output #4
10^-5
1
2
3
4

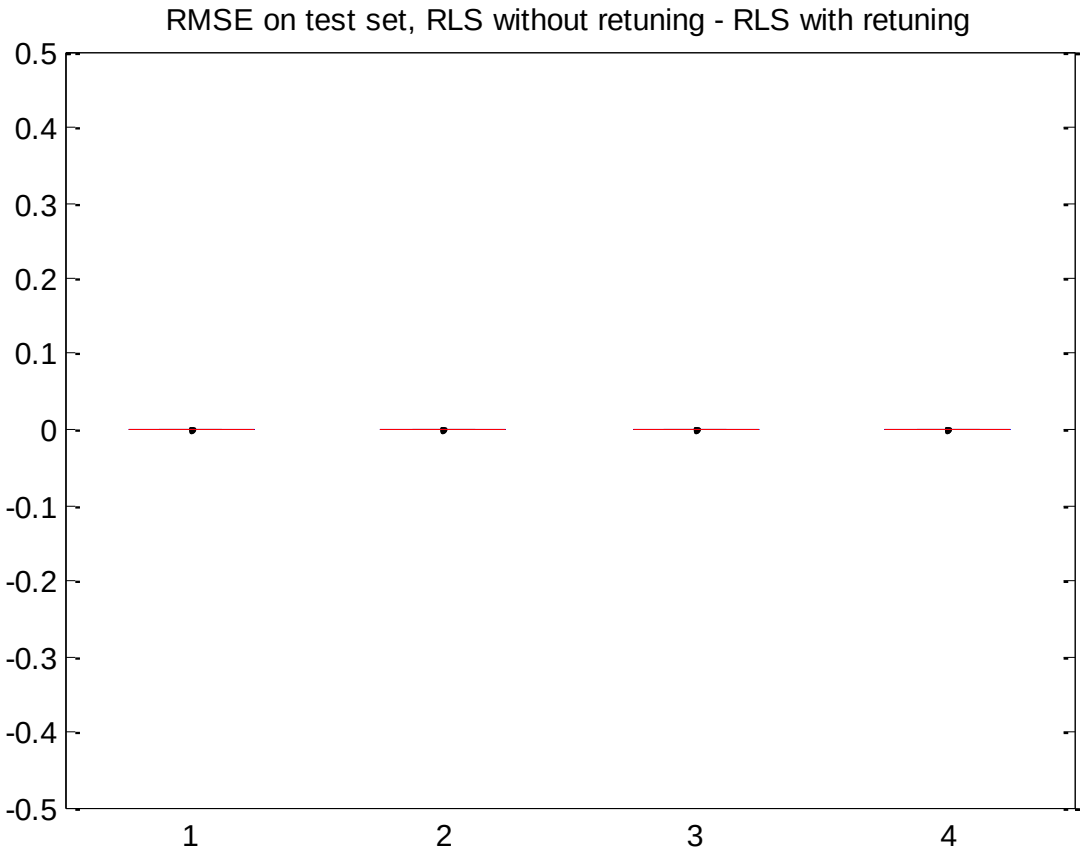

Selected lambdas, RLS with retuning, output #5
10^-5
1
2
3
4

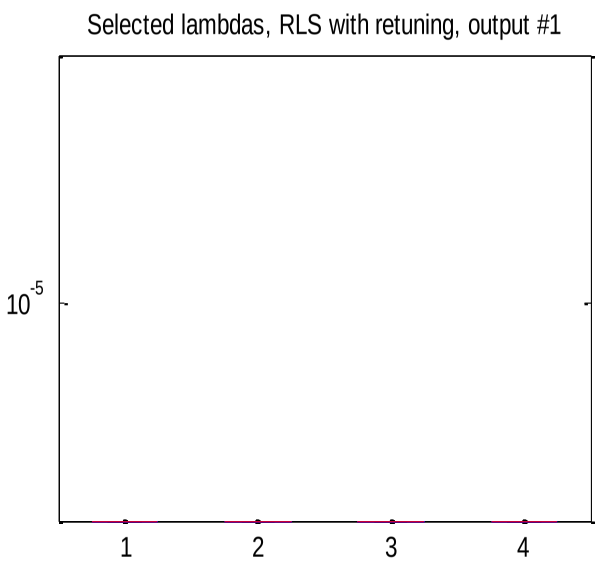

Selected lambdas, RLS with retuning, output #6
10^-5
1
2
3
4

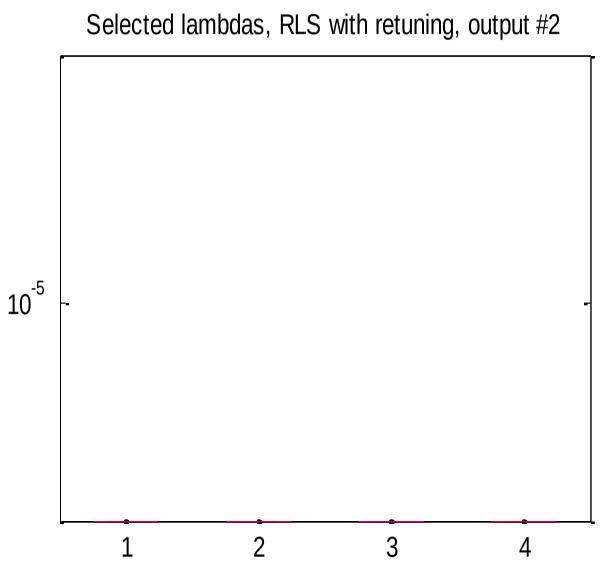

Selected lambdas, RLS without retuning, output #1
Selected lambdas, RLS without retuning, output #2
Selected lambdas, RLS without retuning, output #3
Selected lambdas, RLS without retuning, output #4
Selected lambdas, RLS without retuning, output #5
Selected lambdas, RLS without retuning, output #6
10^-5
1
2
3
4

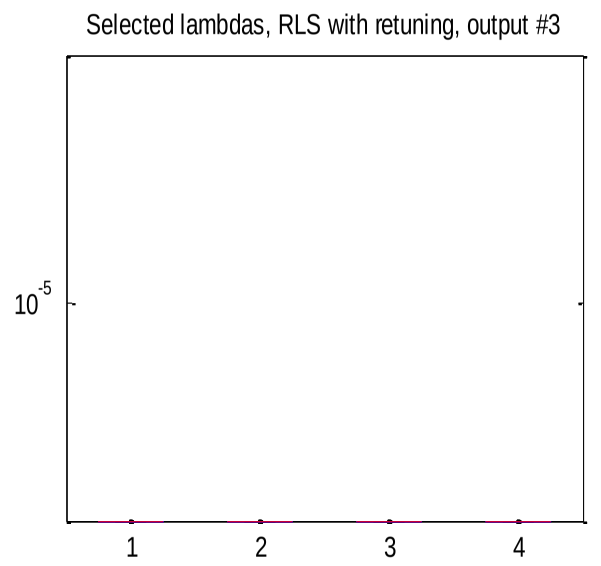
Scaling 10

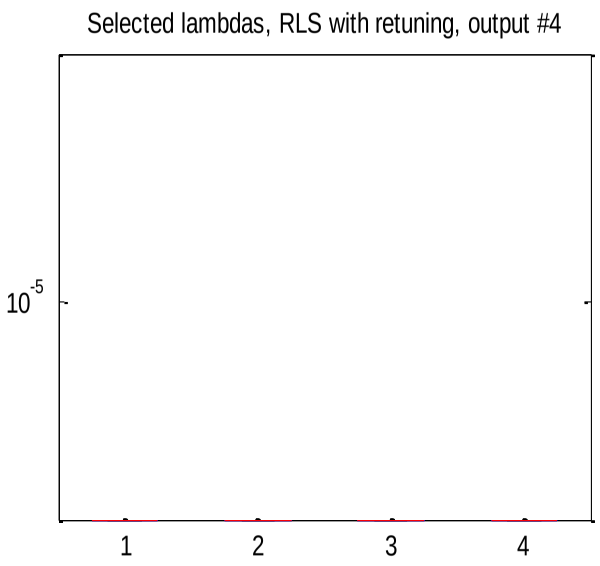

RMSE on test set, RLS with retuning
0.0155
0.015
0.0145
0.014
0.0135
0.013
0.0125
0.012
0.0115
1
2
3
4
RMSE on test set, RLS without retuning
0.0155
0.015
0.0145
0.014
0.0135
0.013
0.0125
0.012
0.0115
1
2
3
4

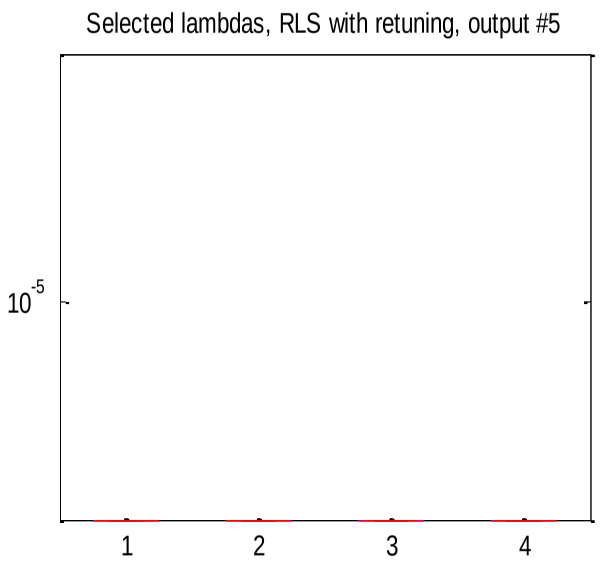

RMSE on test set, RLS without retuning - RLS with retuning
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
1
2
3
4

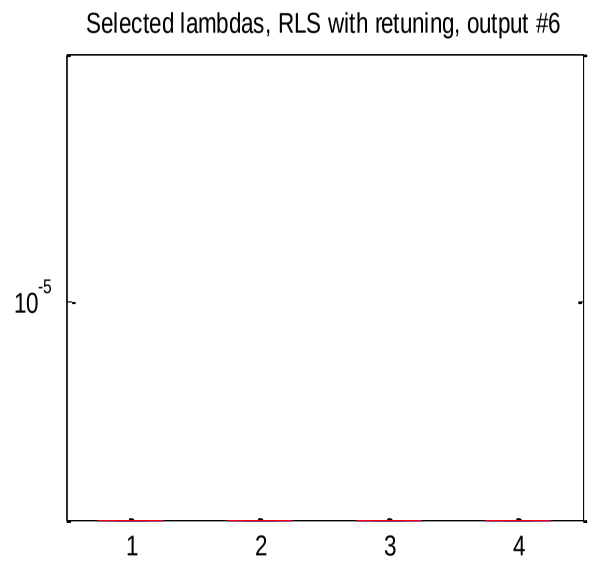

Selected lambdas, RLS with retuning, output #1
10^-5
1
2
3
4

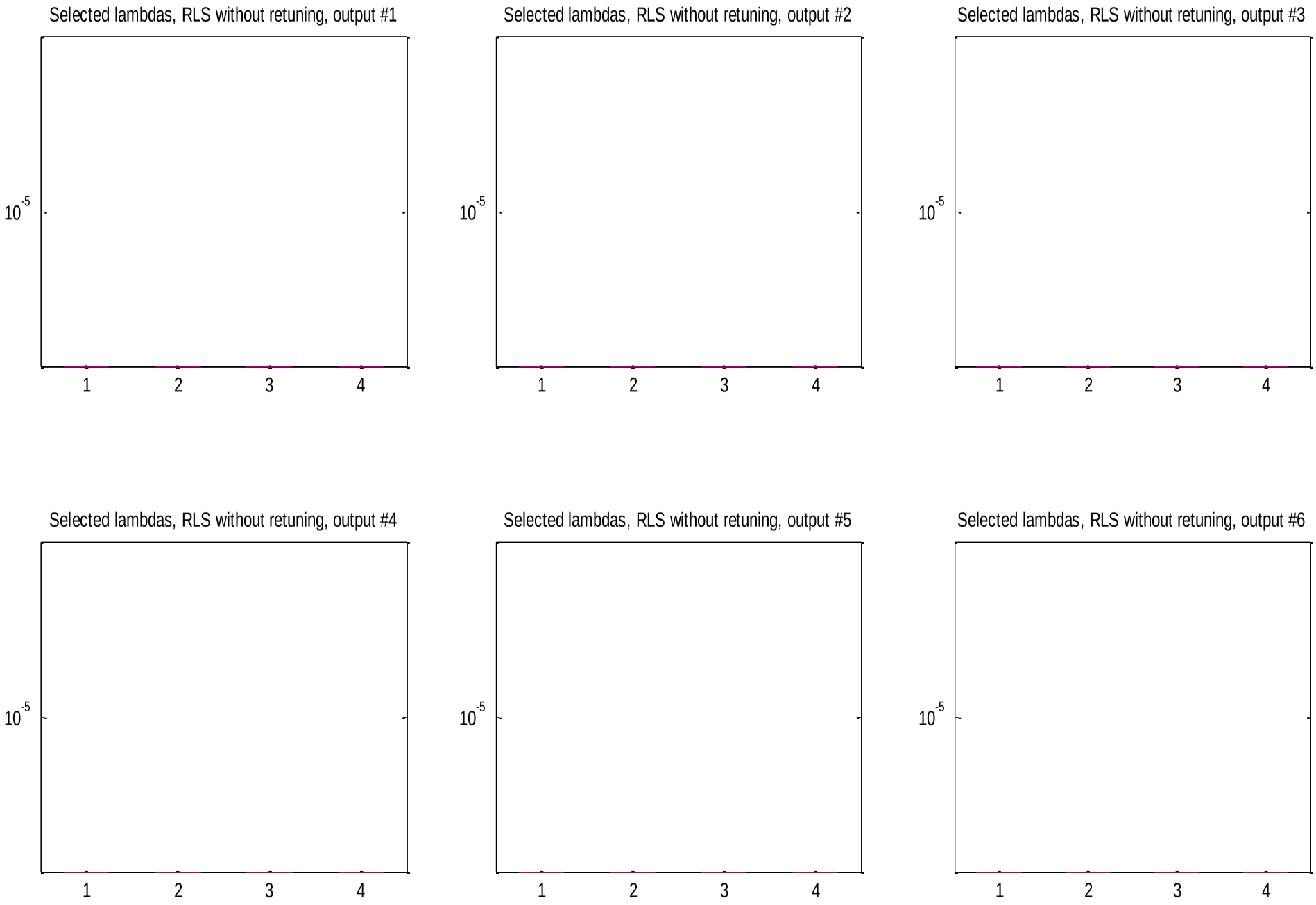

Selected lambdas, RLS with retuning, output #2
10^-5
1
2
3
4


Selected lambdas, RLS with retuning, output #3
10^-5
1
2
3
4

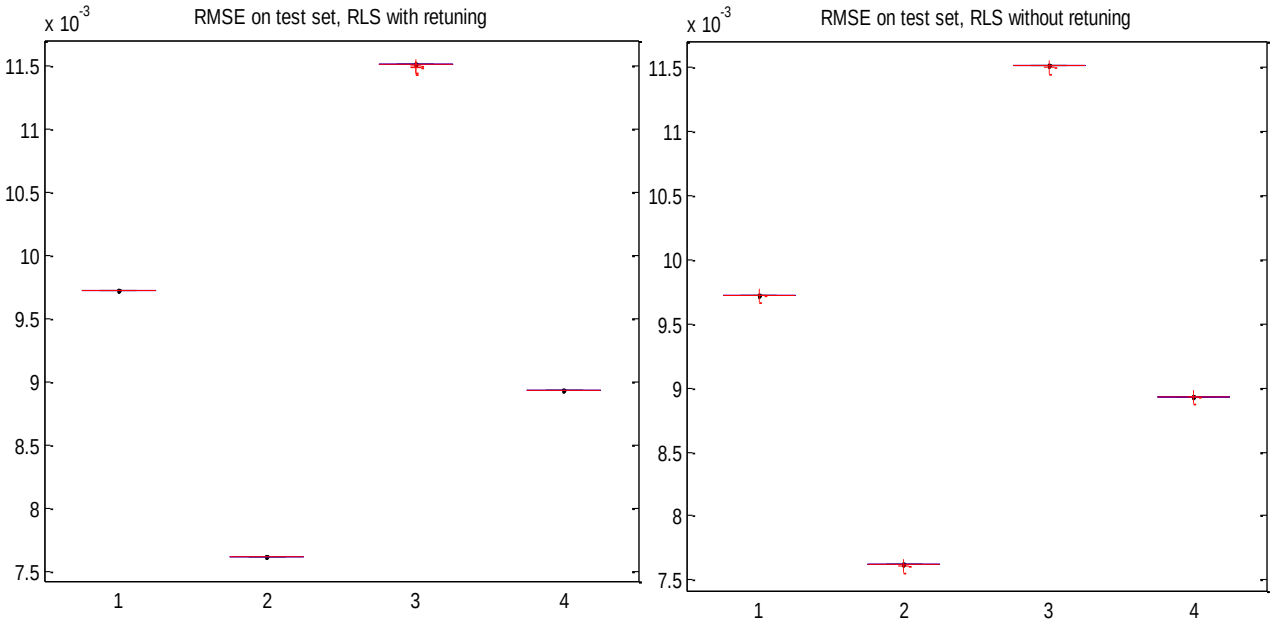

Selected lambdas, RLS with retuning, output #4
10^-5
1
2
3
4

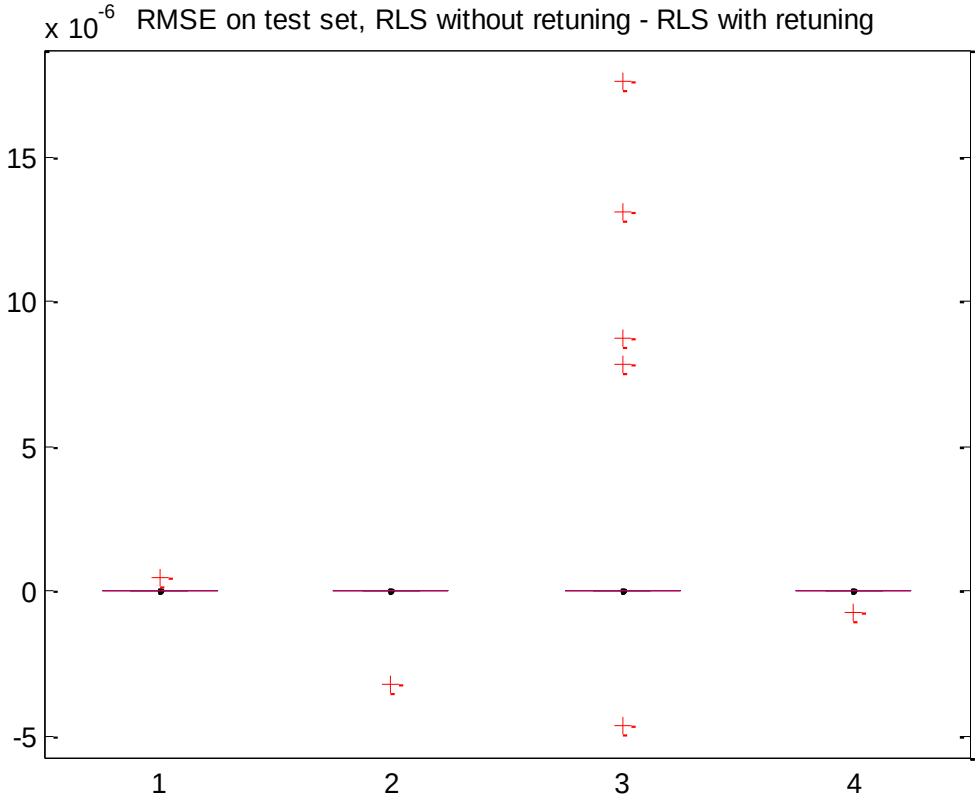

Selected lambdas, RLS with retuning, output #5
10^-5
1
2
3
4

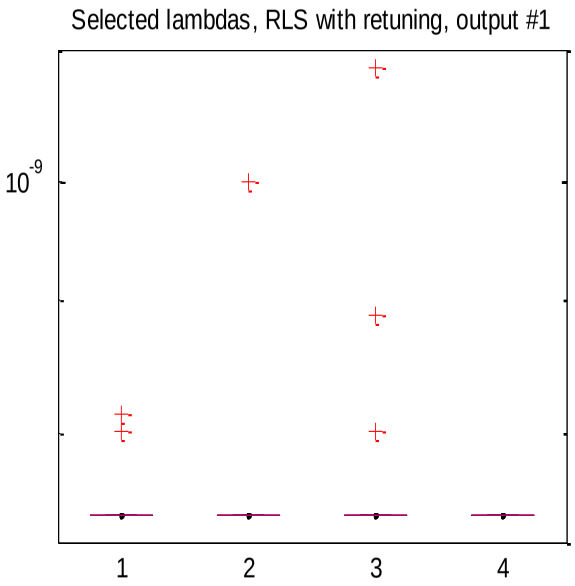

Selected lambdas, RLS with retuning, output #6
10^-5
1
2
3
4

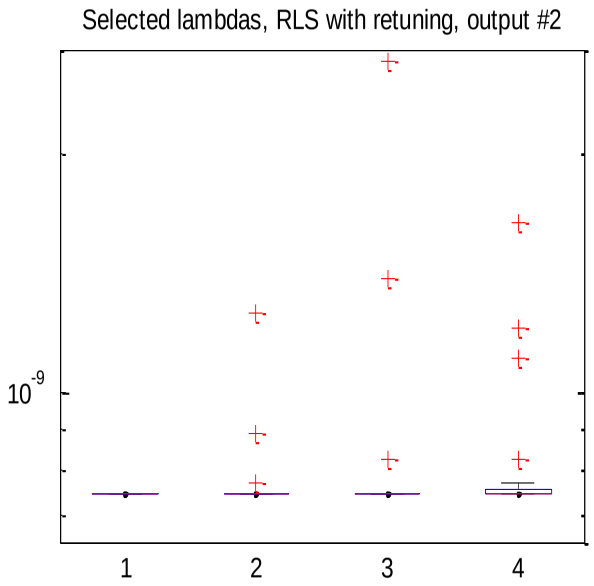

Selected lambdas, RLS without retuning, output #1
Selected lambdas, RLS without retuning, output #2
Selected lambdas, RLS without retuning, output #3
Selected lambdas, RLS without retuning, output #4
Selected lambdas, RLS without retuning, output #5
Selected lambdas, RLS without retuning, output #6
10^-5
1
2
3
4

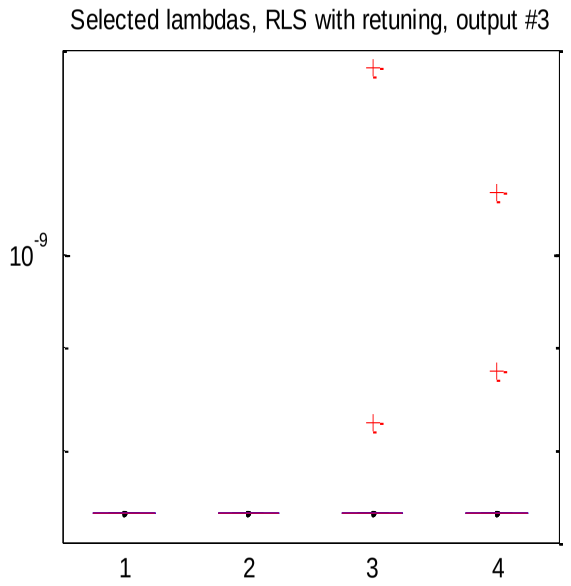
Scaling 100

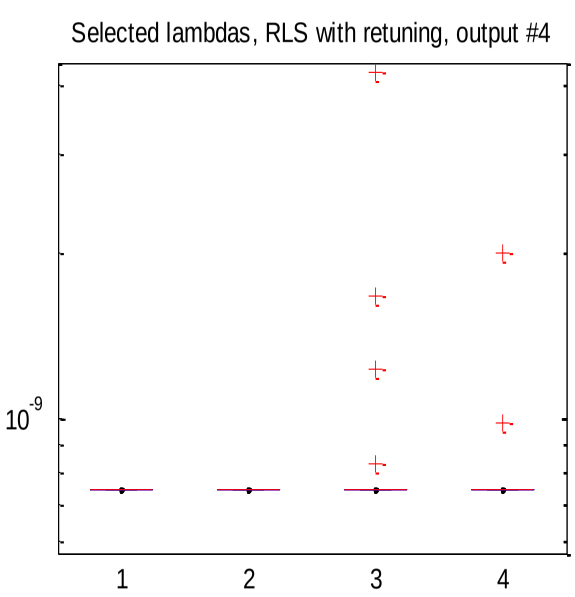

x 10^-3
RMSE on test set, RLS with retuning
11.5
11
10.5
10
9.5
9
8.5
8
7.5
1
2
3
4
x 10^-3
RMSE on test set, RLS without retuning
11.5
11
10.5
10
9.5
9
8.5
8
7.5
1
2
3
4

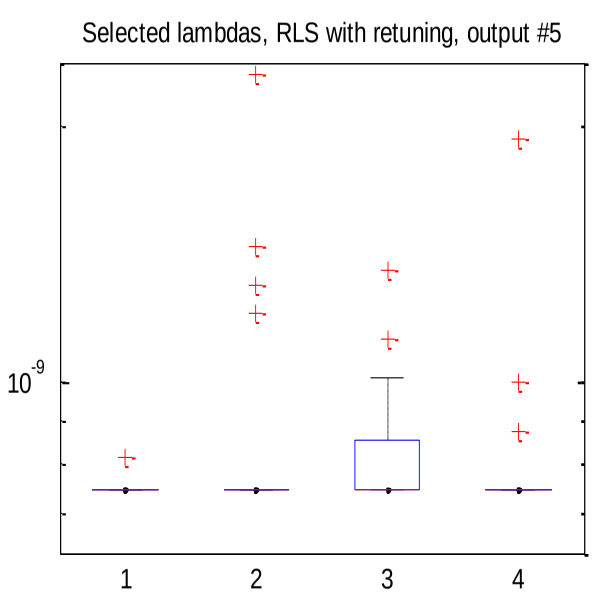

x 10^-6
RMSE on test set, RLS without retuning - RLS with retuning
15
10
5
0
-5
1
2
3
4

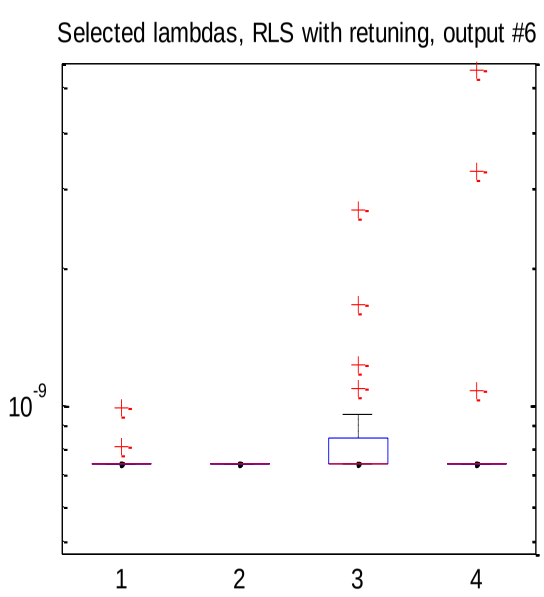

Selected lambdas, RLS with retuning, output #1
10^-9
1
2
3
4

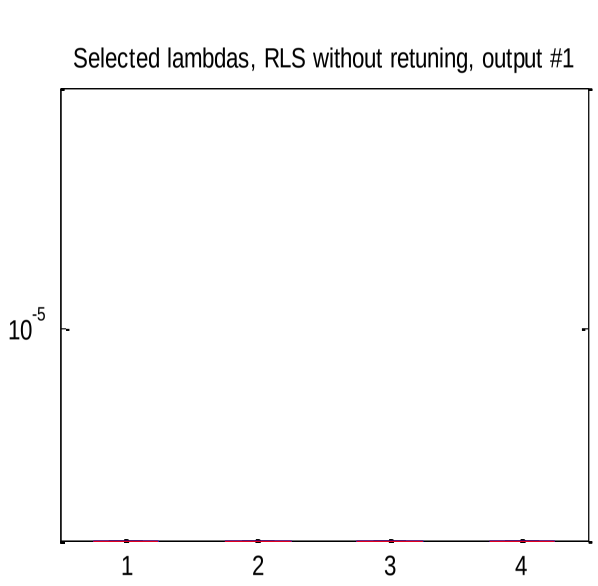

Selected lambdas, RLS with retuning, output #2
10^-9
1
2
3
4

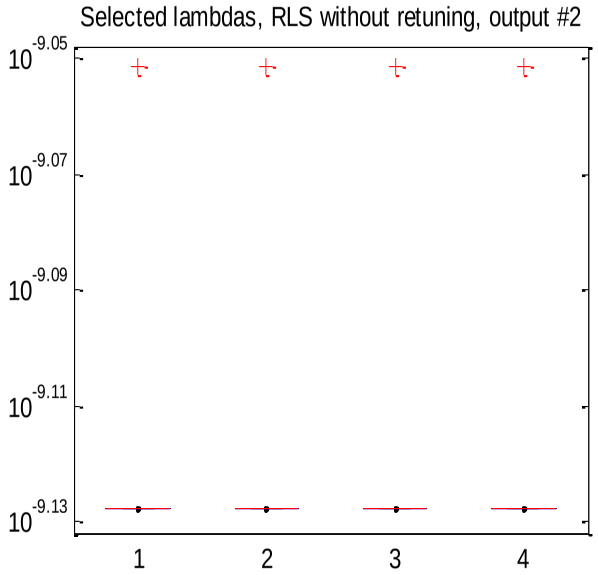

Selected lambdas, RLS with retuning, output #3
10^-9
1
2
3
4

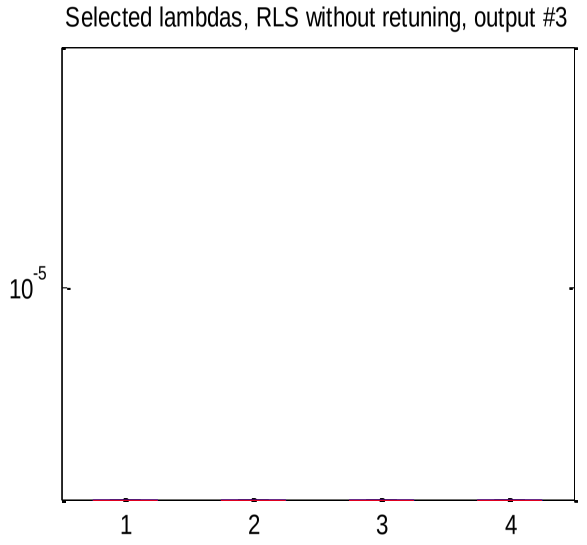

Selected lambdas, RLS with retuning, output #4
10^-9
1
2
3
4

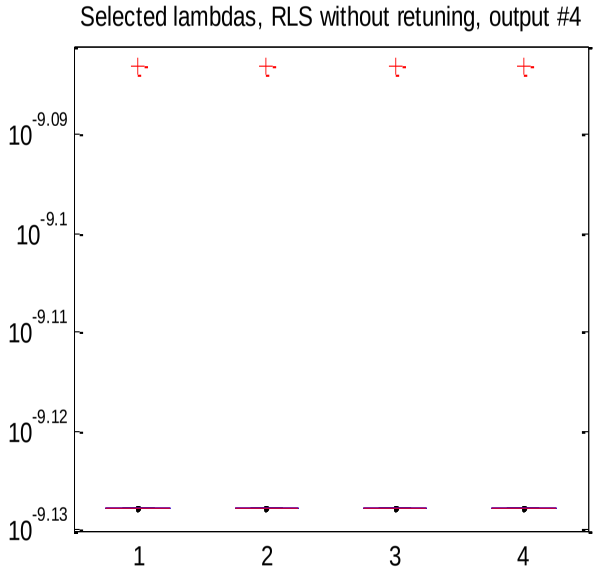

Selected lambdas, RLS with retuning, output #5
10^-9
1
2
3
4

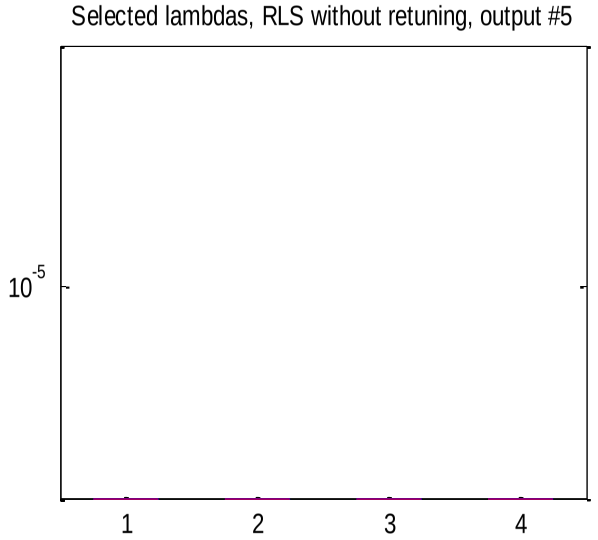

Selected lambdas, RLS with retuning, output #6
10^-9
1
2
3
4

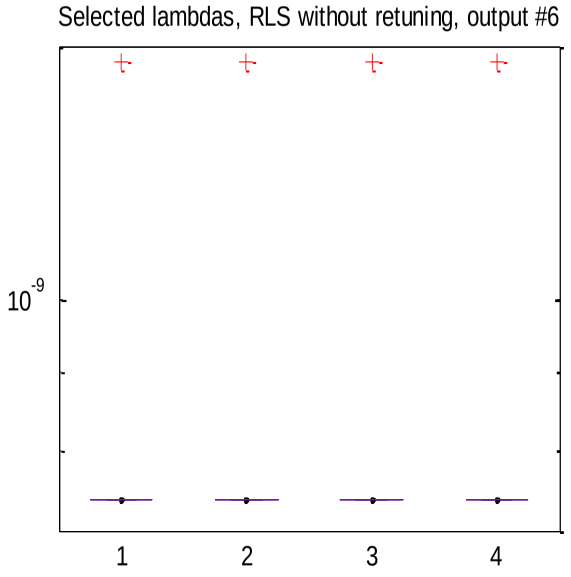

Selected lambdas, RLS without retuning, output #1
10^-5
1
2
3
4


Selected lambdas, RLS without retuning, output #2
10^-9.05
10^-9.07
10^-9.09
10^-9.11
10^-9.13
1
2
3
4

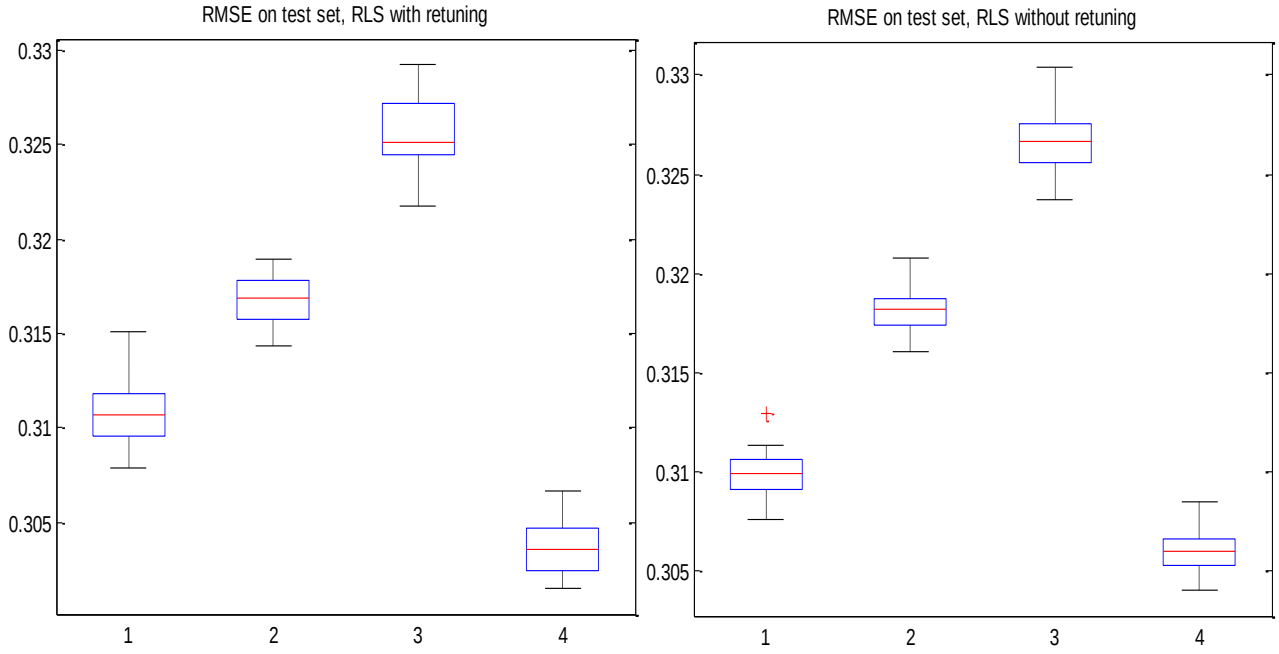

Selected lambdas, RLS without retuning, output #3
10^-5
1
2
3
4

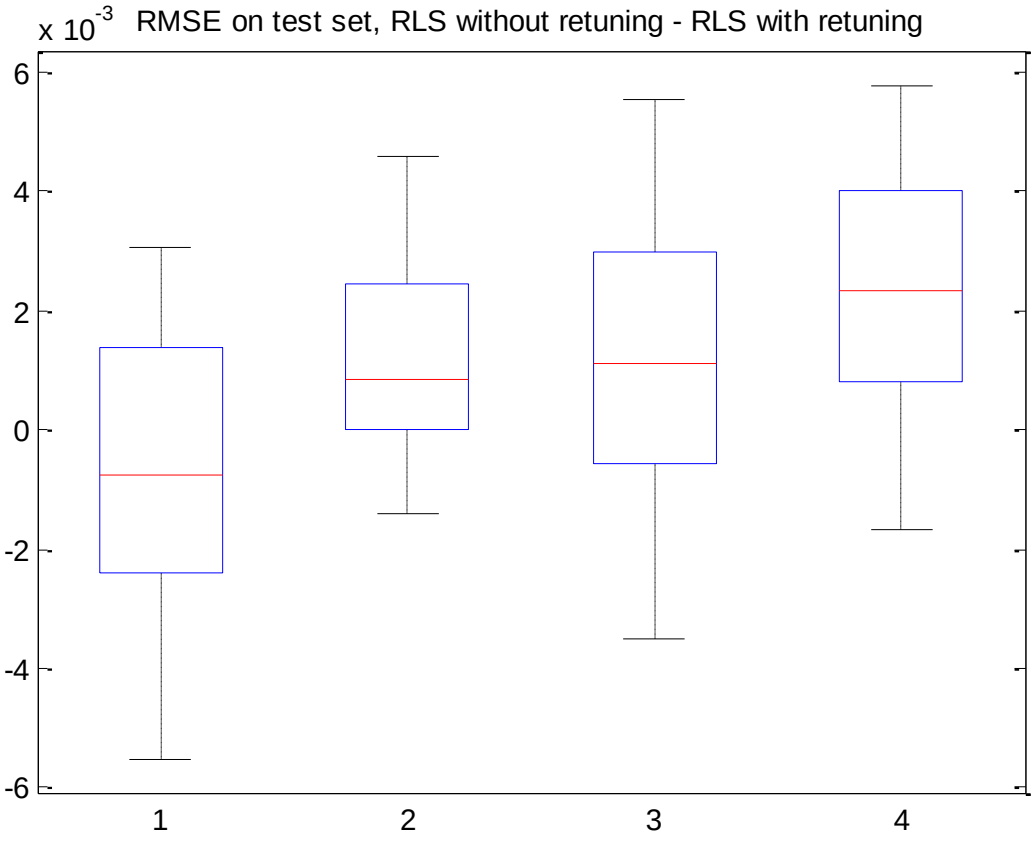

Selected lambdas, RLS without retuning, output #4
10^-9.09
10^-9.1
10^-9.11
10^-9.12
10^-9.13
1
2
3
4

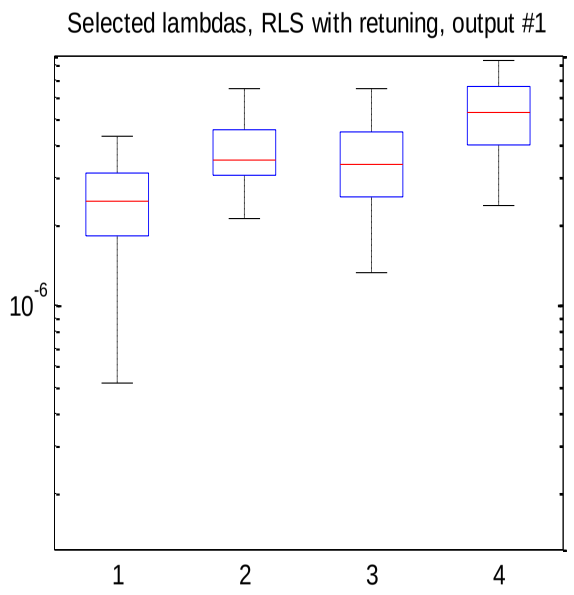

Selected lambdas, RLS without retuning, output #5
10^-5
1
2
3
4

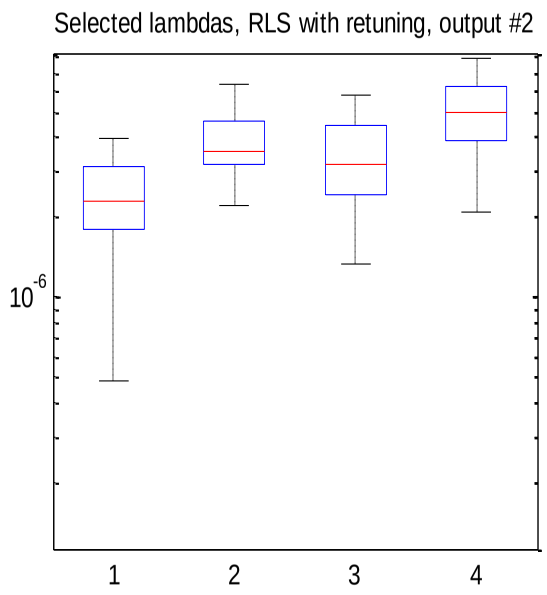

Selected lambdas, RLS without retuning, output #6
10^-9
1
2
3
4

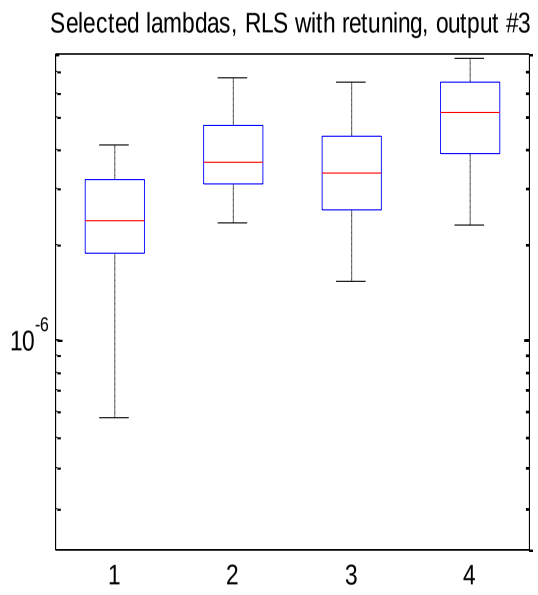
Scaling 1000

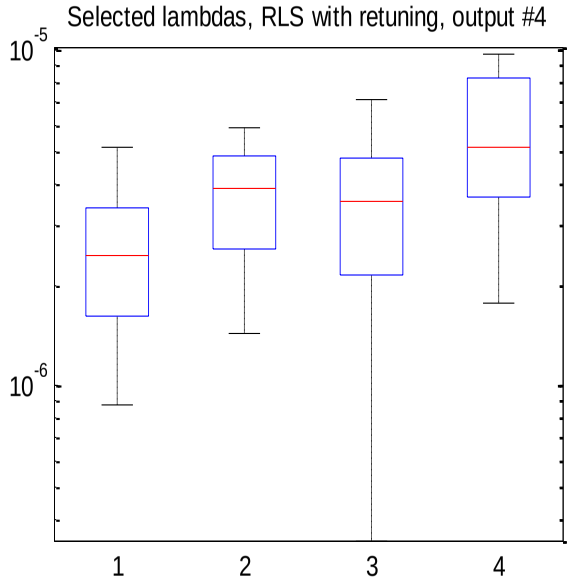


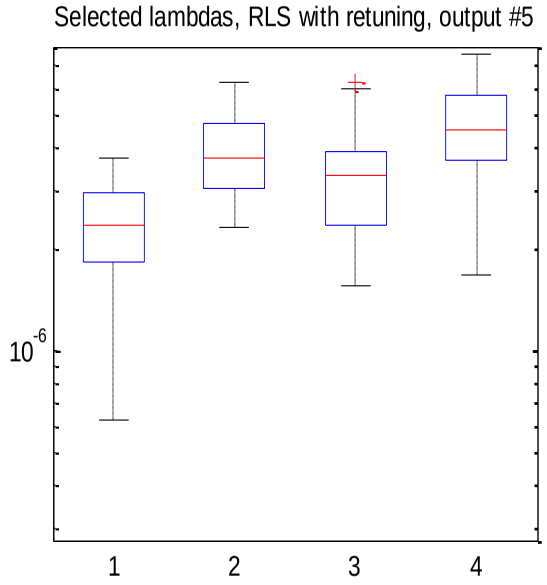


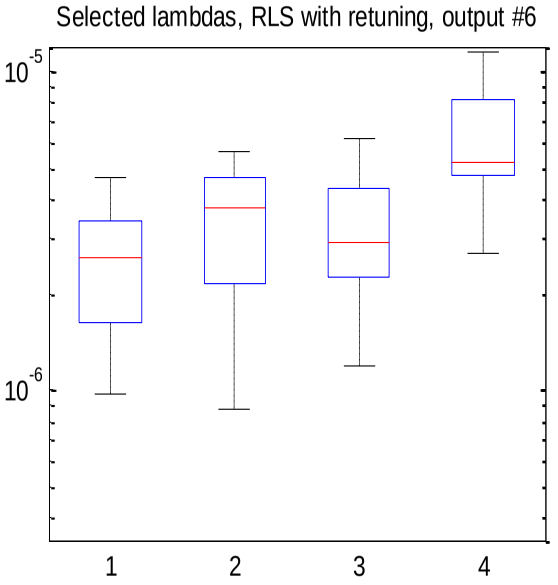

Selected lambdas, RLS with retuning, output #1
10^-6
1
2
3
4

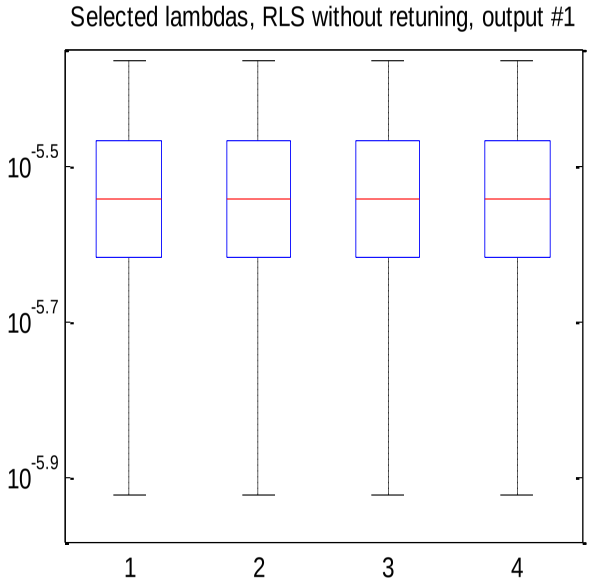

Selected lambdas, RLS with retuning, output #2
10^-6
1
2
3
4

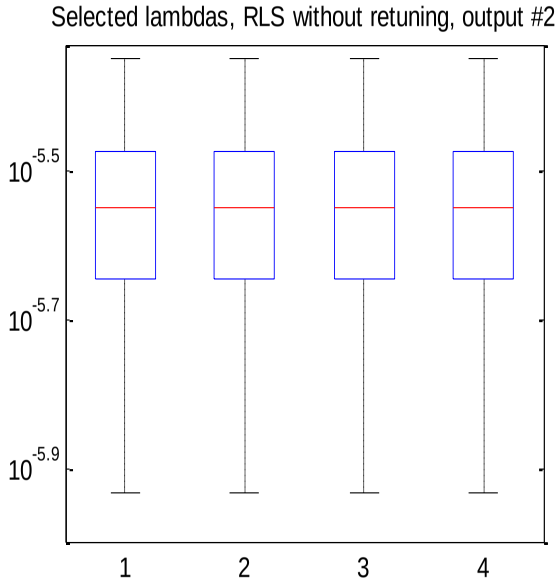

Selected lambdas, RLS with retuning, output #3
10^-6
1
2
3
4

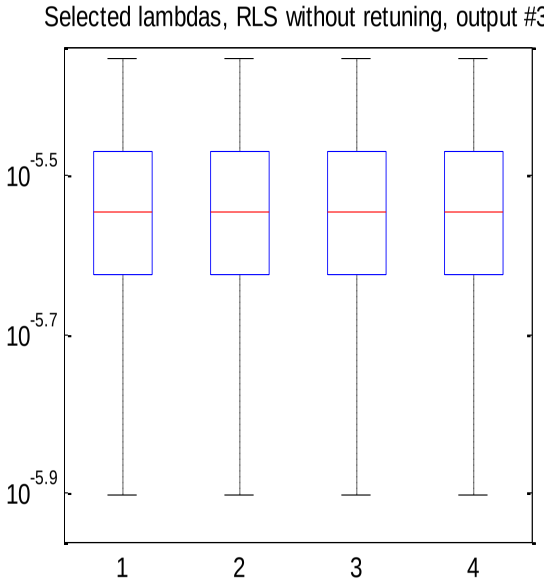

Selected lambdas, RLS with retuning, output #4
10^-5
10^-6
1
2
3
4

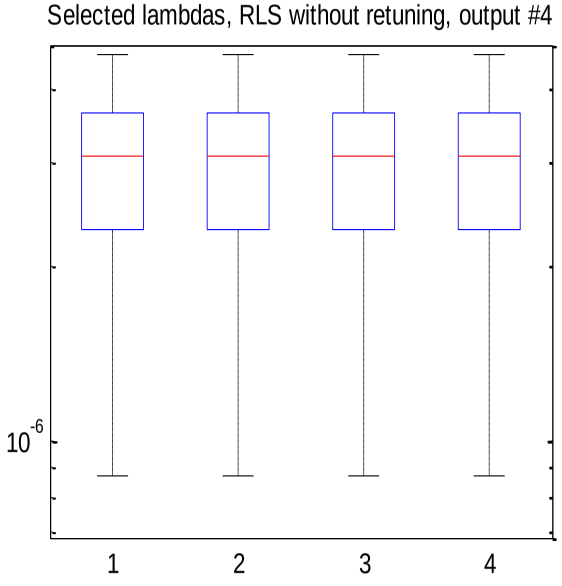

Selected lambdas, RLS with retuning, output #5
10^-6
1
2
3
4

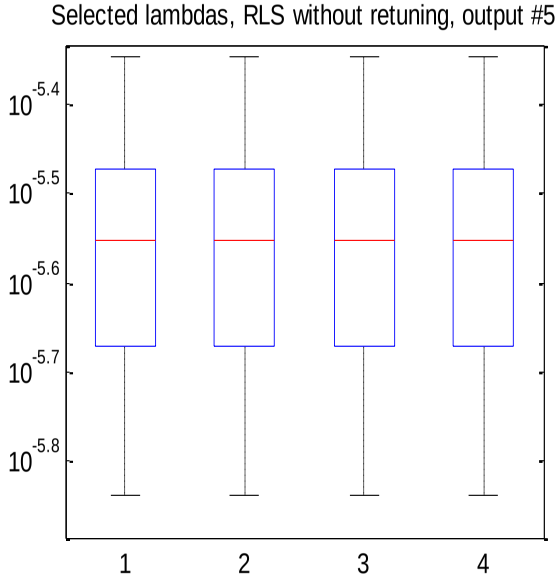

Selected lambdas, RLS with retuning, output #6
10^-5
10^-6
1
2
3
4

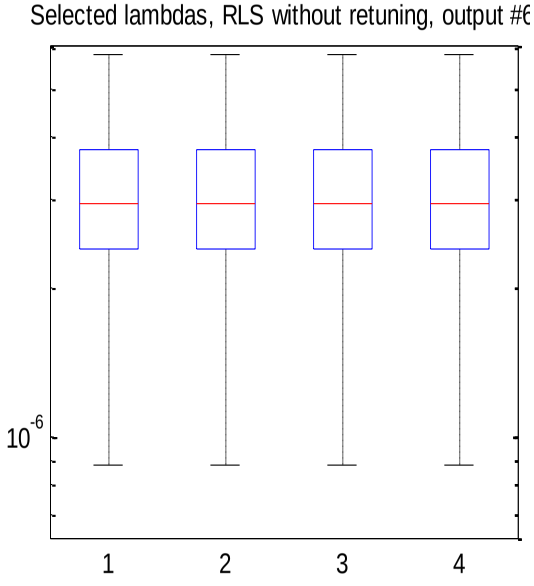

Selected lambdas, RLS without retuning, output #1
10^-5.5
10^-5.7
10^-5.9
1
2
3
4


Selected lambdas, RLS without retuning, output #2
10^-5.5
10^-5.7
10^-5.9
1
2
3
4


Selected lambdas, RLS without retuning, output #3
10^-5.5
10^-5.7
10^-5.9
1
2
3
4


Selected lambdas, RLS without retuning, output #4
10^-6
1
2
3
4


Selected lambdas, RLS without retuning, output #5
10^-5.4
10^-5.5
10^-5.6
10^-5.7
10^-5.8
1
2
3
4


Selected lambdas, RLS without retuning, output #6
10^-6
1
2
3
4

Discussion

Accuracy

- The RMSE is in the order of 10e-2 for *scaling = 1, 10, 100*
- For *scaling = 1000*, the RMSE increases by a factor 10. This phenomenon might be due to the reduction of the width of the kernel gaussians with respect to the scaled distance between the samples (in fact, the selected lambda is higher than in the 1, 10, 100 case, probably for supplying to the missing regularizing effect of sigma).

Regularization

- For *scaling = 1, 10, 100*, the optimal regularization parameter λ is in the order of 10e-9.
- For *scaling = 1000*, λ~10e-6.

Conclusions and future work

- The effect of lambda has been confirmed to be negligible. The reason for this is that the regularizing effect of the random features mapping is already sufficient to reduce the error associated to variance.
- Further investigation on the role of sigma could be fruitful.