HW3: Regularization

Marko Ivanovski

Introduction

For the following task we were supposed to implement Kernelized Ridge Regression and Support Vector Regression. Then we have received two datasets, one which showcased how the methods performed visually and another one on which we had to experiment. Since SVR uses quadratic optimizer which can struggle with large numbers, I first standardized both datasets and proceeded with the analysis.

Sine dataset

The sine dataset can be very well fitted by all combinations. Figure 1 contains the data points, fits, support vectors and margins for SVR. When using the polynomial kernel the parameter 10 gives the best fit, while for the RBF smaller sigma is better. Also in the SVR case, I was able to produce a good sparse fit which does not contain a lot of support vectors by increasing the epsilon (margin).

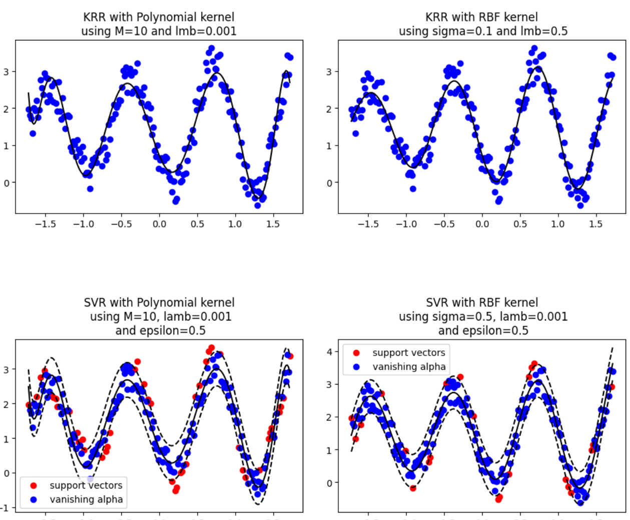

Figure 1. KRR and SVR with Polynomial and RBF kernel applied on the sine dataset. Plots contain the points from the dataset, the fits, the support vectors and margins for the SVR models.

Housing dataset

The housing dataset is more challenging to work with. To this dataset I applied all four combinations of models and kernels. Figure 2 illustrates my final result for different parameters. Every subplot contains one line for RMSE scores on fixed λ 1. The other line is the RMSE when tuned λ was used.

The parameter tuning for λ is essentially a 30 repetitions of 5 K-Fold cross-validation. The tuning was performed only on the training data (first 80%) and the scores are on the test data. For every cross-validation fold I tested different λ values and saved the best one. Finally, I ended up with 5 λs per CV, so 150 λs in total. The collection of λs which I tried fitting models was: [0.01, 0.1, 1, 3, 5, 10, 15, 20, 40, 60, 100]. Our first observation is that the orange (tuned) line is always bellow the blue one which means that the tuning worked well and we fitted the data better.

For the polynomial kernel I am only illustrating values from 1-8 since after that the RMSE on the non-tuned line explodes and it messes the plot ratios. On the RBF kernel I noticed that smaller values performed better which Is the reason I tried smaller values. Finally, I was using a constant ε of 0.01 on the SVR models.

Both KRR and SVR managed to get low RMSE score. The downside of SVR is lower RMSE score results in less sparse fit (a lot of support vectors). Going with smaller ε would have resulted in more sparse solution. Also, since KRR has closed form solution it is much faster to fit, no optimization is needed. But SVR is quite quicker for predictions, once the model is trained since only the support vectors are important for the final prediction.

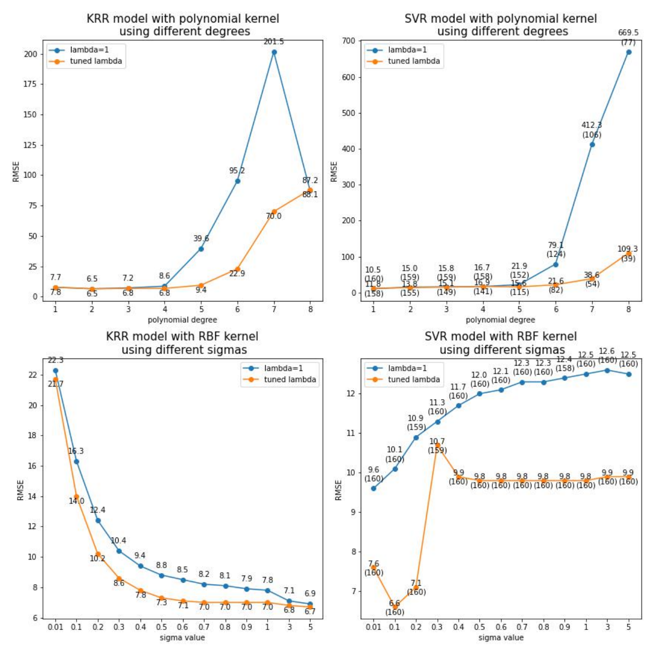

Figure 2. KRR and SVR with Polynomial and RBF kernel applied on the housing dataset. Every subplot then contains a parameter value on x-axis and RMSE score on the y-axis. Also, every subplot contains two lines, one uses fixed λ 1, while the other line shows RMSE on tuned λ. The annotations are the RMSE score and the number of support vectors is in the brackets.