Kernels

Matej Miočić, 63180206, mm9520@student.uni-lj.si

I. Introduction

In this homework we were tasked with implementation of kernelized ridge regression and support vector regression using polynomial kernel and RBF kernel. First we show parameter influence on a simple sine dataset for each method and kernel pair. Then we try to select the best lambda (λ) parameter using cross validation on housing2r dataset and compare RMSE with models fit with $\lambda = 1$.

II. Models

A. Parameter influence

Each method has a regularization parameter lambda (λ). In this section we select $\lambda = 0.01$, since we are using the whole sine dataset to fit the model and we allow it to overfit for explanation purposes. In Figure 1 we show fit of both methods using polynomial kernel with degrees $\mathrm{M} \in \{1, 5, 10\}$. Since support vector regression requires another parameter, epsilon (ϵ), we set it to $\epsilon = 0.5$, to nicely fit the sine data (see Figure 3 for more on ϵ parameter).

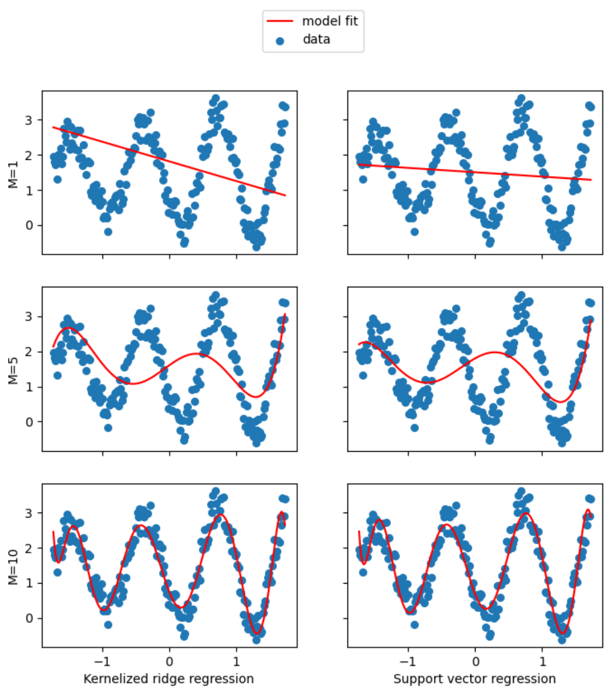

Figure 1. Fit of kernelized ridge regression (left) and support vector regression (right) on sine dataset with **polynomial** kernel with degrees $\mathrm{M} \in \{1, 5, 10\}$.

In Figure 2 we show fit of both methods using RBF kernel with $\sigma \in \{0.1, 0.5, 1\}$. In both Figures (1 and 2) we observe minimal differences between methods. Both methods seem to fit best with polynomial degree $\mathrm{M} = 10$ using polynomial kernel and parameter $\sigma = 0.5$ using RBF kernel.

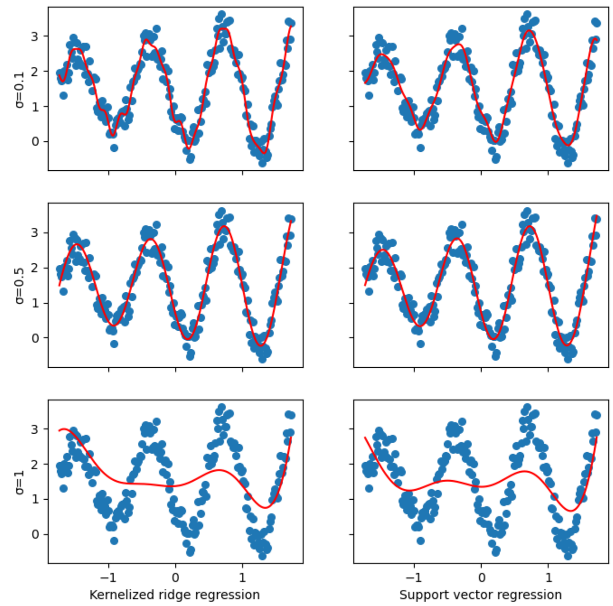

Figure 2. Fit of kernelized ridge regression (left) and support vector regression (right) on sine dataset with **RBF** kernel with $\sigma \in \{0.1, 0.5, 1\}$.

In Figure 3 we show how ϵ parameter affects SVM. Instances outside the ϵ-tube are allowed to have α, α^* coefficients more than 0. These instances with the nonvanishing coefficients (non-zero) are called support vectors. Using lower values squeezes the ϵ-tube, making the fit more precise and increases the number of support vectors.

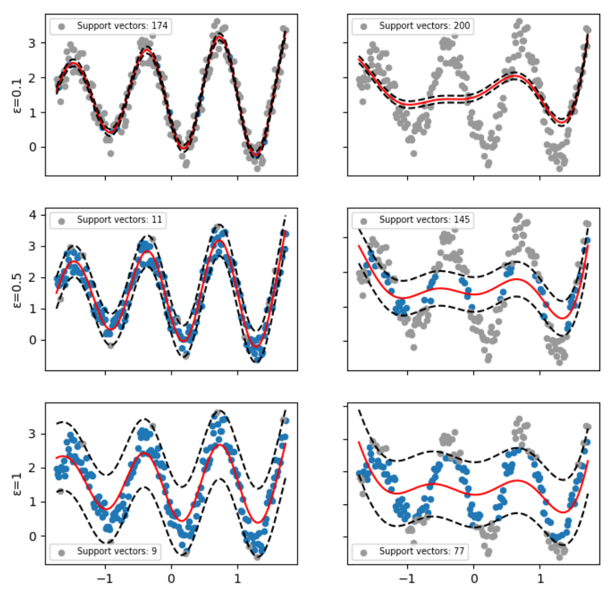

Figure 3. Support vector regression fit (red line), ϵ-tube (black dotted lines) and support vectors (gray dots) for $\epsilon \in \{0.1, 0.5, 1\}$ using RBF kernel on sine dataset with $\sigma = 0.5$ (left) and $\sigma = 1$ (right).

B. Predicting with methods

In this section we show how methods perform with predicting the housing2r data set. We took first 160 instances as the training set and set the rest 40 instances as the test set. We standardized the dataset using information only from the training set. In Figures 4 and 5 we plot RMSE for predicted test set using regularization parameter $\lambda = 1$ and λ obtained by performing 5-fold cross validation 10 times with different folds. For SVM we also display the number of support vectors. We experimented with ϵ parameter to minimize the number of support vectors while still getting a good RMSE score. For this dataset we decided to use $\epsilon = 5$. For polynomial kernel we plot results for $M = \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, while for RBF we decided to plot results for $\sigma \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$.

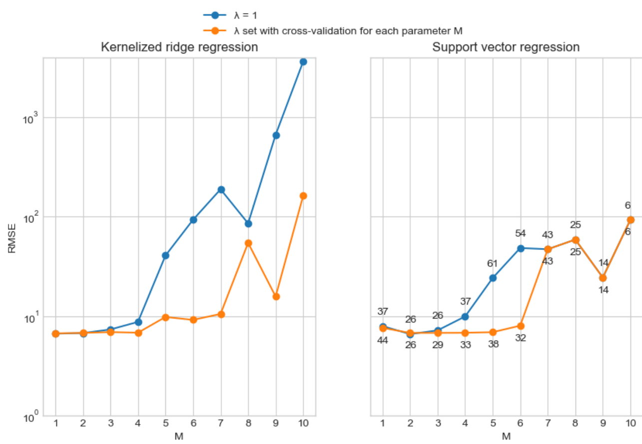

Figure 4. Log-scale of RMSE for prediction of test set with kernelized ridge regression (left) and support vector regression (right) using **polynomial** kernel for parameters $M = \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, with $\lambda = 1$ (blue line) and tuned λ (orange line). For SVM we also display number of support vectors (for $\lambda = 1$ above and tuned λ below).

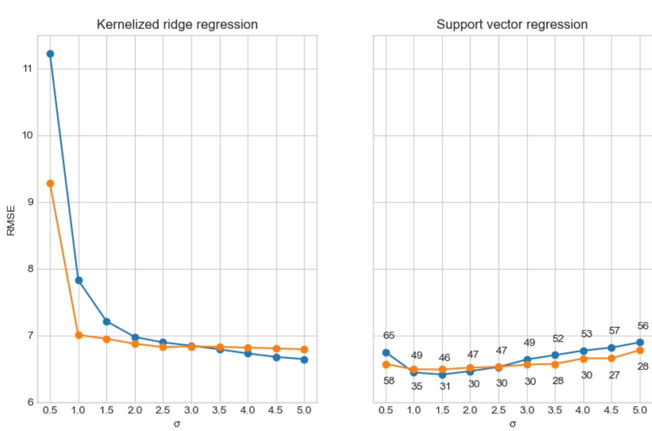

Figure 5. RMSE for prediction of test set with kernelized ridge regression (left) and support vector regression (right) using **RBF** kernel for parameters $\sigma \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$, with $\lambda = 1$ (blue line) and tuned λ (orange line). For SVM we also display number of support vectors (for $\lambda = 1$ above and tuned λ below).

We observe a lower RMSE with tuned λ when using polynomial kernel, meanwhile we see that RMSE does not change by much when using RBF kernel. Our explanation is that we are already getting good results with $\lambda = 1$, that tuning the parameter could not help by much. However we do observe a lower number of support vectors when using tuned λ parameter. Overall we obtained better results with support vector regression than with kernelized ridge regression.

III. Conclusion

We learned how to implement kernelized ridge regression and support vector regression. We showed that support vector regression performs a bit better on the housing2r dataset at the cost of computation speed. Our conclusion is that if we want to predict the data faster we should use support vector regression. However if we want to fit our data faster we should use kernelized ridge regression.