Homework 5: Kernels

Stefanela Stevanović
MLDS1 23/24, FRI, UL
63220492

I. Introduction

The goal of this homework was to implement and investigate the performance of two popular kernel-based regression methods: kernelized ridge regression (KRR) and Support Vector Regression (SVR), with a focus on two commonly used kernels: the polynomial kernel and the RBF kernel. We applied these regression methods and kernels to a 1-dimensional sine data set and the housing2r data set, a real-world regression problem. By evaluating the performance of these regression methods and kernels with different parameters, on both synthetic and real-world data, we provide insights into their strengths and weaknesses for different regression tasks.

II. Experiments

A. Experiment on sine data set

In this experiment, we applied two regression methods, namely kernelized ridge regression and Support Vector Regression (SVR), to a 1-dimensional sine dataset standardized to the mean of 0 and a standard deviation of 1. For both kernelized ridge regression and SVR, we experimented with two types of kernels: the polynomial kernel and the RBF kernel. Our goal was to find suitable kernel and regularization parameters that yield accurate fits to the data. For SVR, we also took care to produce a sparse solution. During the experimentation phase, we explored different values for the kernel and regularization parameters. For the polynomial kernel, we varied the parameter M, which determines the degree of the polynomial. As for the RBF kernel, we experimented with different values of the parameter sigma, which controls the spread of the kernel function. The kernelized ridge regression fits are shown in Fig.1, while SVR fits with polynomial and RBF kernel are shown in Fig.2 and Fig.3, respectively.

B. Experiment on housing2r data set

The housing2r data set was divided into a training set and a validation set using an 80-20 split. The first 80% of the data were assigned to the training set, while the remaining 20% were allocated to the validation set. The data was standardized by transforming the features of both the training and validation sets to have a mean of 0 and a standard deviation of 1. For each kernel and regression method combination, we varied the kernel parameter, evaluated the model+kernel combination on the validation set and plotted results for the which we get the lowest MSE on the validation set. For the polynomial kernel, a range of polynomial degrees (M) from 1 to 10 was considered, while for the RBF kernel σ values ranging from 1 to 10 were explored. For each kernel and regression method combination, two curves were plotted: one with a fixed regularization parameter λ = 1 and varying kernel parameter, and the other with the optimal λ determined for each kernel parameter separately. Obtained results are presented in Fig.4 and Fig.5.

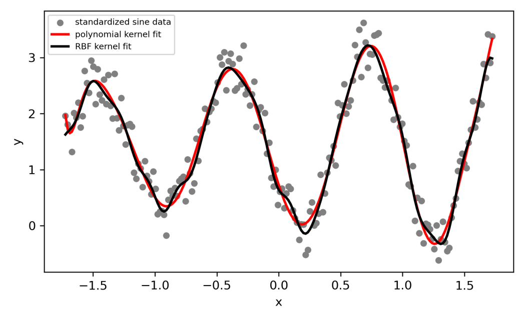

Fig. 1. Kernelized ridge regression fits for the standardized sine data using both the polynomial (red) and RBF (black) kernels. Polynomial fit was achieved with parameters λ = 0.2 and M = 15, while RBF fit was achieved with parameters λ = 0.2, σ= 0.1. Kernelized ridge regression performs well on the sine dataset using both kernels. The polynomial kernel achieves MSE of 0.075, while the RBF kernel achieves MSE of 0.065. However, upon closer examination, we observe that the polynomial kernel better captures the sinusoidal behavior, resulting in a smoother curve. On the other hand, the RBF tends to slightly fit the noise in the data.

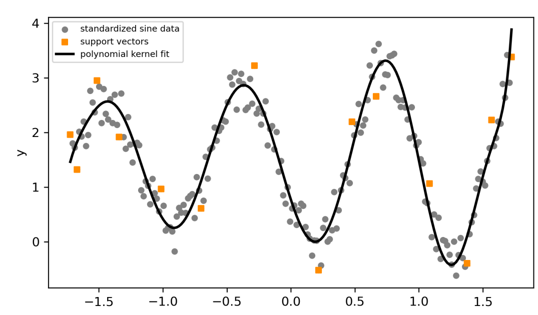

Fig. 2. Support vector regression with polynomial kernel fit for the standardized sine data. By setting the parameters to λ = 0.4, M = 15, and ϵ = 0.5, we were able to obtained a good fit with MSE of 0.085, while also producing a sparse solution. The model identified 14 support vectors, which are visually represented by orange squares.

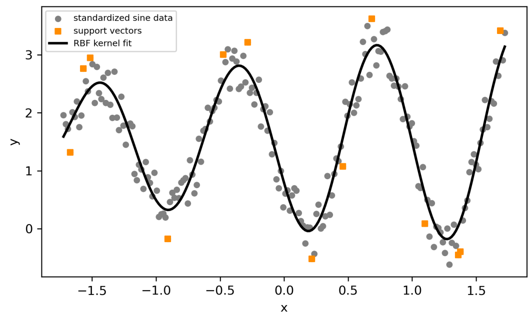

Fig. 3. Support vector regression with RBF kernel fit for the standardized sine data. With parameters $\lambda = 0.3$, $\sigma = 0.3$, and $\epsilon = 0.5$, we achieved a good fit with MSE of 0.080 and a sparse solution with 13 support vectors. Identified support vectors are visually represented by orange squares.

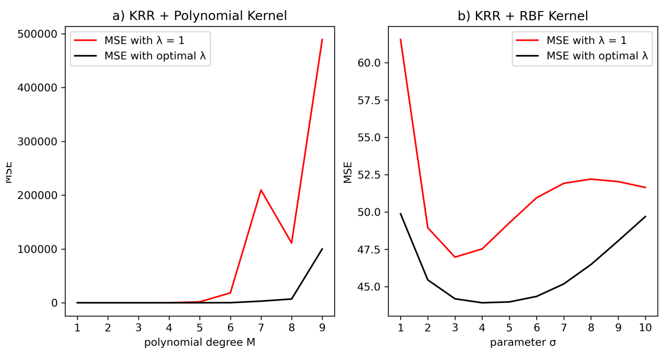

Fig. 4. Kernelized ridge regression MSE on the housing2r validation set. a) KRR with polynomial kernel: The MSE is plotted for varying polynomial degree M using a constant λ value of 1 (red) and optimal λ values for each M (black). The MSE value for M = 10 has been omitted from the plot due to its exceptionally large value. b) KRR with RBF kernel: The MSE is plotted for varying σ values using a constant λ value of 1 (red) and optimal λ values for each σ (black). The figure clearly demonstrates the importance of selecting appropriate regularization parameters, as indicated by the superior performance achieved with the optimal λ values compared to using a constant value of 1, for both polynomial and RBF kernel.

III. Conclusion

From our experimenting with different combinations of regression models and kernels, we observed that the choice of kernel had a greater impact on the model's performance than the choice of regression model itself. The type of kernel determines the model's ability to capture different types of patterns, highlighting the importance of selecting the appropriate kernel. In second experiment, when we varied the kernel parameters and determined the optimal lambda for each parameter, we found that with the RBF kernel, the optimal λ value was generally low and remained relatively stable across varying σ values on out housing2r data set. On the other hand, with the polynomial kernel, the optimal λ value exhibited significant changes as the polynomial degree varied and was significantly higher, suggesting the need for stricter regularization when using the polynomial kernel compared to the RBF. This makes the RBF kernel more robust, in a sense that model's performance remains stable and less dependent on the choice of the kernel and regularization parameter.

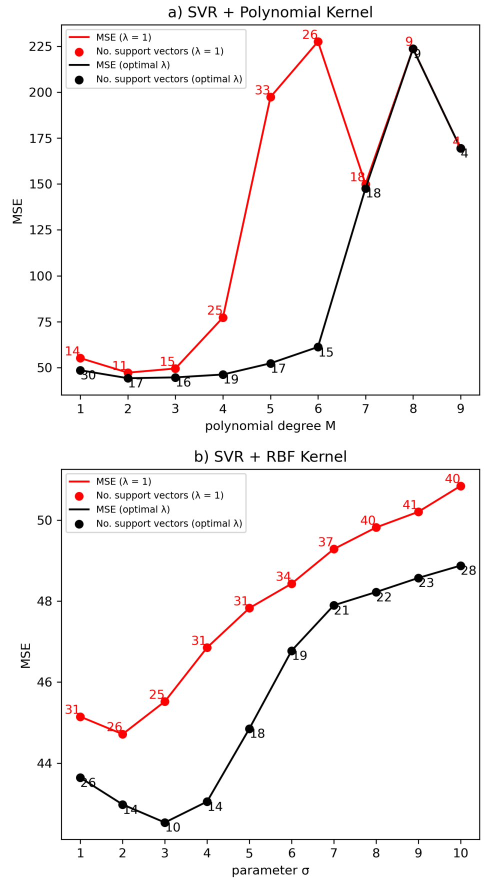

Fig. 5. Support vector regression MSE on the housing2r validation set. The parameter ϵ was set to 8, based on its optimality in achieving a good fit and promoting model sparsity. a) SVR with polynomial kernel: The MSE is plotted for varying polynomial degree M using a constant λ value of 1 (red) and optimal λ values for each M (black). b) SVR with RBF kernel: The MSE is plotted for varying σ values using a constant λ value of 1 (red) and optimal λ values for each σ (black). The number of support vectors for each kernel parameter is displayed on the graph for both the $\lambda = 1$ and optimal λ cases. The inclusion of the number of support vectors helps visualize the impact of the kernel parameters on the model complexity. We observe that choosing the optimal λ values not only improves the MSE, resulting in lower errors, but also usually leads to solutions with a reduced number of support vectors, indicating a more sparse solution. This is especially true for the SVR with RBF kernel.

In terms of the model comparison between the KRR and SVR, SVR also demonstrates more robustness, both in terms of the kernel type used and the kernel parameter variation. The changes in kernel parameters result in smaller changes in MSE on the validation set with SVR compared to KRR. As for the choice of the most suitable combination of regression model and kernel for our housing2r data set, we believe that SVR with the RBF kernel would be the best choice. The reasons for this have already been explained, but it is worth mentioning that SVR with the RBF kernel demonstrated the lowest MSE on our data set with the parameter values: $\sigma = 3$, $\lambda = 0.1$, and $\epsilon = 8$. This parameter combination also resulted in a sparse solution, with only 10 support vectors used.