HW5: Kernels

Žan Pušenjak
MLDS 2024/25 , FRI, UL
zp4613@student.uni-lj.si

***Abstract*—We implemented kernelized regression models and tested them on a provided data sets. We were tasked with finding the best set of parameters for the data. We used `nested-cross-validation` to compare the performance at different parameters.**

I. Implementation

We implemented the **kernelized ridge regression** (KRR) and **Support Vector Regression** (SVR). To go along with the models we implemented three different kernels:

- Polynomial kernel
- RBF kernel
- Linear kernel

We used the implementations on sine data. We could observe that the choice of the parameters greatly affected the fit. With SVR we had to be careful to manage the number of *support vectors*. We used parameters in TableI and were able to achieve results seen on Figure6

TABLE I: Parameters for KRR and SVR on sine data

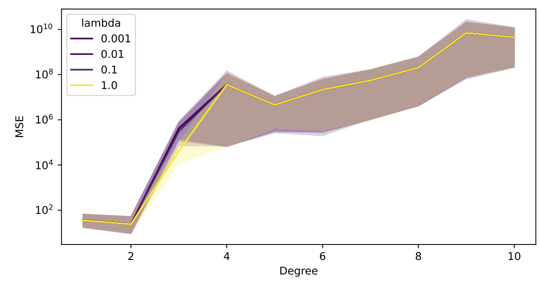
Model	Kernel	Parameters	
KRR	RBF	$\lambda = 1$, $\sigma = 0.5$	$\epsilon = 1$
KRR	Polynomial	$\lambda = 0.001$, $M = 1$	$\epsilon = 0.4$, $c = 1$
SVR	RBF	$\lambda = 1$, $\sigma = 0.5$	
SVR	Polynomial	$\lambda = 0.1$, $M = 1$	$c = 1$

Where M is the degree of the polynomial and c is the coefficient in the polynomial.

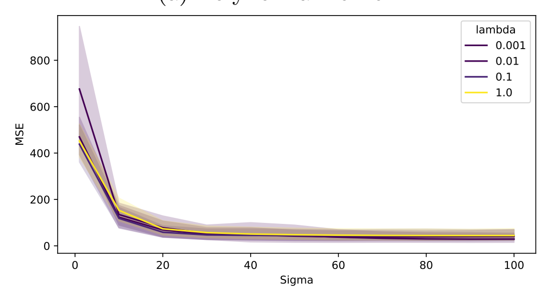
We observed that the `RBF` kernel achieved an excellent fit to the data, while the `Polynomial` kernel struggled to capture the underlying pattern effectively.

II. Fitting the housing data

We trained our model to fit the housing data. The combinations of the kernels and methods stayed the same, but this time we used `nested-cross-validation` (NCV) to determine the best parameter λ. We also fixed $\lambda = 1$ and compared the two results. To determine the ϵ parameter for SVR, we made tried to fit the model on some samples and looked at the number of the `support vectors`. Observing the results in II we can see that the parameter did not change the number so we fixed it at $\epsilon = 0.5$



(a) Polynomial kernel



(b) RBF kernel

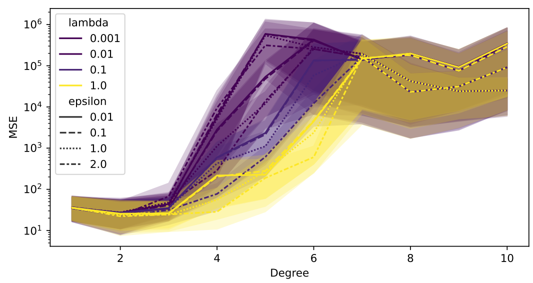
Fig. 1: KRR parameter exploration

TABLE II: Effect of ϵ on MSE and number of support vectors

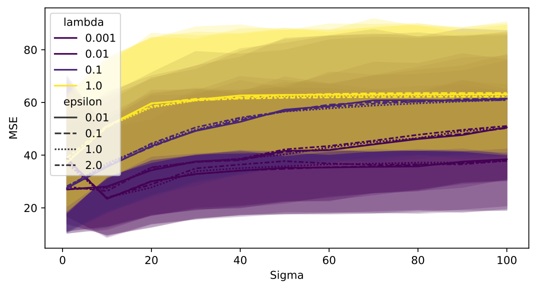
ϵ	MSE	Support Vectors
0.1	26.53 ± 12.05	95.89 ± 28.47
0.3	26.47 ± 12.05	91.42 ± 28.44
0.5	26.43 ± 12.07	96.17 ± 24.60
0.7	26.34 ± 12.03	98.69 ± 25.83
0.9	26.26 ± 12.01	101.30 ± 23.25

Because computation of lots of parameter values was not possible we evaluated the models with a simple train-test split to get a feel for what values of parameters to explore more deeply. The results can be observed on Figure 1 for KRR and Figure 2 for SVR.

Thanks to this simple exploration we were able to confidently chose parameters $M \in [1, 10]$, $\sigma \in \{80, 100, 120\}$ and $\lambda \in \{0.01, 0.1, 1, 10\}$. We started with `leave-one-out NCV` but because of computational reasons we abandoned that approach and used a `10-fold NCV`.

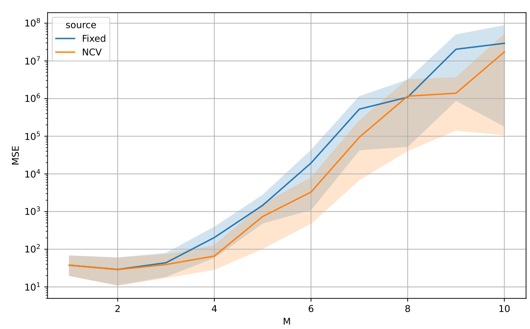


(a) Polynomial kernel

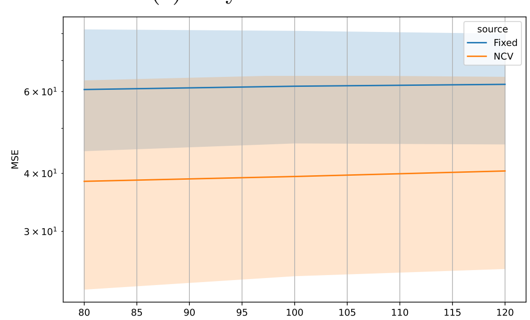


(b) RBF kernel

Fig. 2: KRR parameter exploration



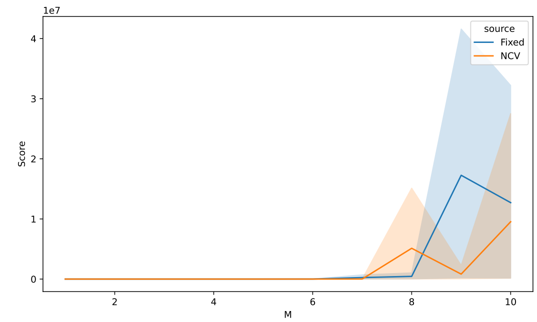
(a) Polynomial kernel



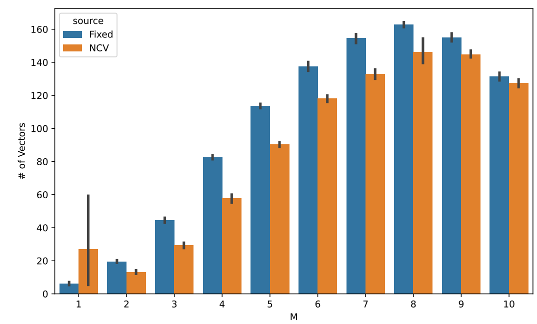
(b) RBF kernel

Fig. 3: KRR

The results can be observed on Figures 3, 4 and 5. We can see that the NCV approach outperforms the fixed λ model, and would likely perform even better if we could check even more values for λ. We can see that with polynomial kernel, the number of support vectors increases as the degree increases where with RBF kernel the sigmas do not have any effect, but the fixed lambda yields less vectors.

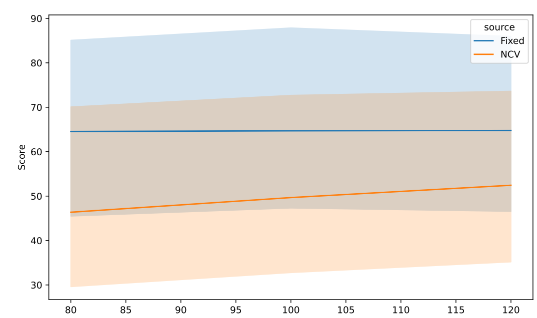


(a) Performance at different parameters

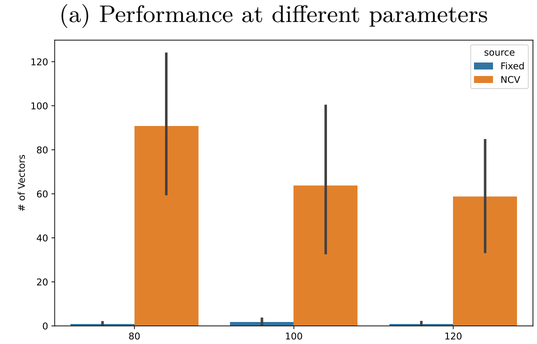


(b) Number of support vectors

Fig. 4: SVR with Polynomial kernel



(a) Performance at different parameters



(b) Number of support vectors

Fig. 5: SVR with RBF kernel

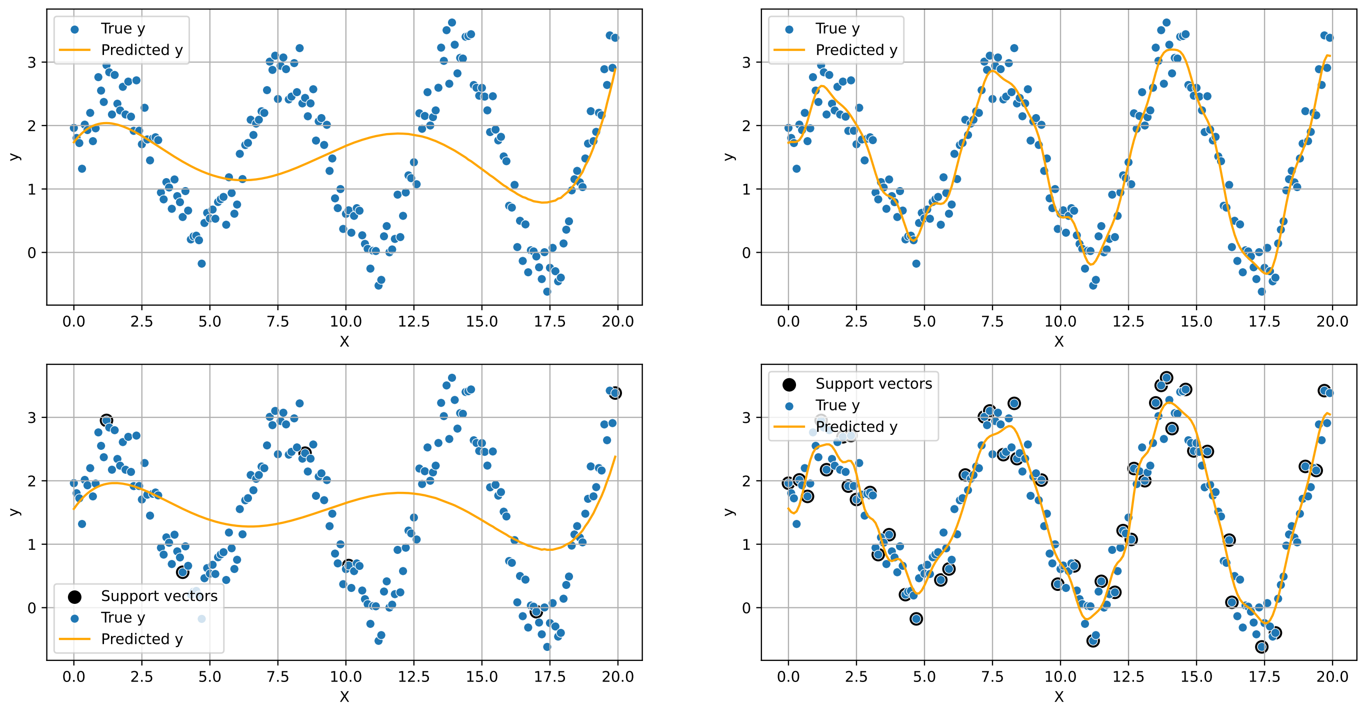


Fig. 6: Plot of fitted regressions on sine data