Support Vector Regression

Žiga Trojer - 63200440

0.1 Sine data

Like in the previous homework, we tested our algorithm on the `sine.csv` data. We used same kernels, but we used support vector regression (SVR) instead of kernelized ridge regression (KRR). The best choice of parameter ε was 0.5. Regularization parameter λ is set to 0.1. Figure 1 shows the predictions of data points with SVR using Polynomial kernel with $M = 11$ and RBF kernel with $\sigma = 0.3$. With this choice of parameters, the solution is sparse. We also marked support vectors on a plot - there are 19 support vectors using Polynomial kernel and 15 support vectors using RBF kernel. Fitting a sine data using SVR is a bit slower than using KRR, but we don't see major improvement compared to it (in terms of how good the fit is).

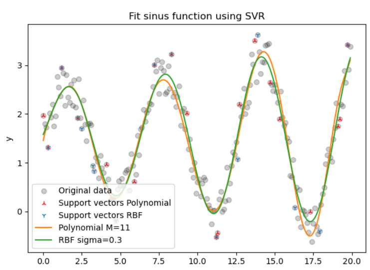

Figure 1. Fitted sine data using support vector regression and both kernels.

0.2 Housing data

We also applied SVR to the *housing* dataset. We have used the same random seed when splitting the data, so we could compare results from our new algorithm with the results from the previous homework. Figure 2 contains of 2 graphs. The graph above shows RMSE on testing set, using polynomial kernel and different choices of parameters λ - first with $\lambda = 1$ and then with optimal λ for each value of M in range $[1, 10]$, which we got from 5-fold cross-validation. For λ we tested values $\{0.001, 0.01, 0.1, 1, 5, 10, 25, 50, 100\}$ and for each M chose the best one. The graph below shows number of support vectors for each choice of parameters. Parameter ε was chosen beforehand - we tested few parameters and chose the best one (in a sense that we have a good fit and minimal number of support vectors). We notice that number of support vectors starts to rise at $M = 3$ until the value $M = 6$ and then it starts to fall again. But with the increase of support vectors, we see also the increase of RMSE, which is interesting. Optimal values of λ are $\{0.1, 25, 100, 100100, 100, 100, 5, 0.001, 0.1\}$. Around $M = 6$ the regularization parameter λ is 100 (best choice) which means that miss-classifications are less allowed than with $\lambda = 1$ (for the other model with constant λ), so the RMSE does not increase so much.

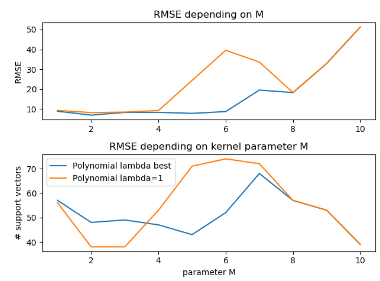

Figure 2. RMSE and number of support vectors depending on chosen parameter value M.

Figure 3 shows RMSE and number of support vectors using RBF kernel. Similar as before, we chose best λ for the choice of σ. We see that we need more support vectors for RBF kernel, but RMSE is lower for each choice of parameter λ - *best* and $\lambda = 1$. Number of support vectors increase when increasing σ too much, so does the RMSE. Because RMSE is lower, we would choose this model when having this data.

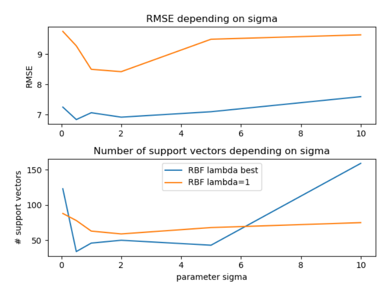

Figure 3. RMSE and number of support vectors depending on chosen parameter value σ.

When comparing results with results from the previous homework, we see that SVR using Polynomial kernel is more powerful than KRR, since RMSE is lower by a lot! We notice some similarities too: RMSE from both algorithms increase with M, but there was no local maximum at the kernelized ridge regression. In both cases RMSE with values from CV is lower than with the predetermined one. If we compare results for the RBF kernel, obvious decrease of RMSE is seen when increasing σ for a bit. But when increasing σ too much, RMSE starts to increase again - both have this property. SVR performed similarly as KRR in terms of RMSE, maybe even a bit worse for the choice of $\lambda = 1$. We would choose KRR in this case, since performance in terms of RMSE is similar as SVM, but the algorithm is faster, which in many cases, have a big importance when choosing algorithm.