MLDS: Homework 5

Andraž De Luisa

April 10, 2020

Implement support vector regression with two kernels: polynomial $\kappa(x, x') = (1 + xx')^M$ and RBF kernel: $\kappa(x, x') = \exp(-\frac{||x-x'||^2}{2\sigma^2})$.

Task 1 Figure 1 shows the fitting of the support vector regression model on the sine dataset, with two different kernels: a polynomial and a RBF kernel. The values of the kernel parameters are not optimal with respect to any measure, I have just chosen such values that visually fit well to the actual sine values from the data. A smaller value of the degree M in the polynomial kernel would result in a flattened curve, while a higher M would cause a clear overfit. A similar behaviour could be seen with the RBF kernel, with the lost of smoothness (i. e. overfit) for lower σ, and a flattening for higher ones. The margin $\epsilon = 0.5$ was chosen in such way to obtain a sparse model (less than 10% of support vectors in this case).
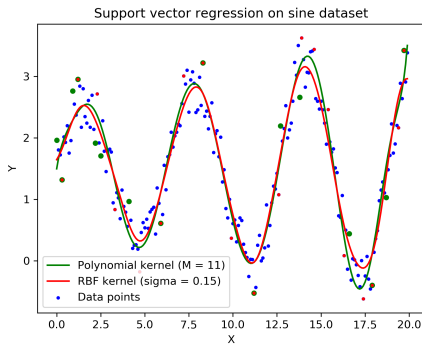

Figure 1: Support vector regression with a polynomial and a RBF kernel, both with regularization parameter $\lambda = 0.01$. Support vectors for each model are marked.

Task 2 Figures 2 and 3 show the evaluation of the support vector regression with polynomial and RBF kernel on the test data from the housing dataset.

As the measure of predictive performance the RMSE (rooted mean squared errors) was used, both in parameter optimization and model evaluation. Both figures present a comparison between a fixed regularization parameter $\lambda = 1$ and the best λ for each kernel parameter (M or σ), set with internal 5-fold cross-validation. In figure 2 just the degrees up to 4 are shown, because the errors grow too much from degree 5 onwards due to overfit. With both kernels the CV estimation of the regularization parameter seem to slightly improve the performance of the model, but we can't really make any conclusion about it because of the uncertainty involved.
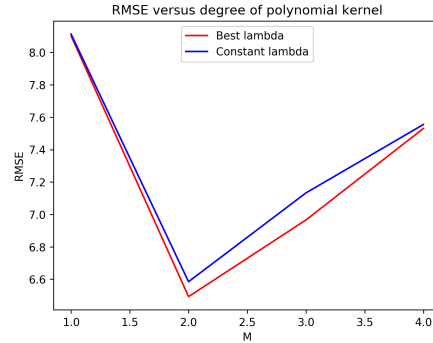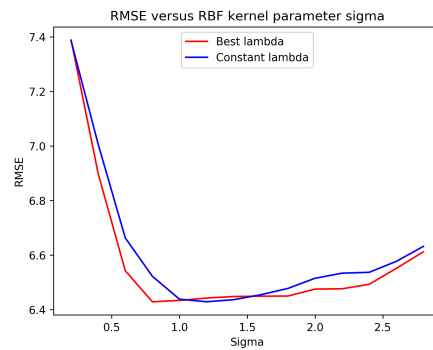

Figure 2: Polynomial kernel

Figure 3: RBF kernel