Homework 4: Kernels

Matej Kalc

Models implementation

Kernelized Ridge Regression

I have implemented the Kernelized Ridge Regression in the close form solution.

$$\beta = X^T(k(X,X^T) + \lambda I_n)^{-1} y \tag{1}$$

$$\hat{y}(x') = \beta^T x' \tag{2}$$

k(x, y) is the selected kernel function. In 1 we fit or train the model. In 2 predict the outcome of instance x'. In b

Support Vector Regression

SVR has been implemented by solving a quadratic problem using the cvxopt library. The quadratic problem is stated in equation 3. The used matrices for fitting the SVR model are shown in equations 6, 4, 8, 5. k(x, y) is the selected kernel function and C = $\frac{1}{\lambda}$.

$$\text{minimize } \frac{1}{2}x^T P x + q^T x \text{ where } Qx \leq h \text{ and } Ax = b \tag{3}$$

$$P = \begin{bmatrix} k(X_1,X_1) & -k(X_1,X_1) & \dots & k(X_1,X_n) & -k(X_1,X_n) \\ -k(X_1,X_1) & k(X_1,X_1) & \dots & -k(X_1,X_n) & k(X_1,X_n) \\ k(X_2,X_1) & -k(X_2,X_1) & \dots & k(X_2,X_n) & -k(X_2,X_n) \\ -k(X_2,X_1) & k(X_2,X_1) & \dots & -k(X_2,X_n) & k(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k(X_n,X_1) & -k(X_n,X_1) & \dots & k(X_n,X_n) & -k(X_n,X_n) \\ -k(X_n,X_1) & k(X_n,X_1) & \dots & -k(X_n,X_n) & k(X_n,X_n) \end{bmatrix} \tag{4}$$

$$q = \begin{bmatrix} \varepsilon - y_1 \\ \varepsilon + y_1 \\ \varepsilon - y_2 \\ \varepsilon + y_2 \\ \vdots \\ \varepsilon - y_n \\ \varepsilon + y_n \end{bmatrix} \qquad h = \begin{bmatrix} C \\ C \\ \vdots \\ C \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{5}$$

$$A = \begin{bmatrix} 1 & -1 & 1 & -1 & \dots & 1 & -1 \end{bmatrix} \tag{6}$$

$$b = \begin{bmatrix} 0 \end{bmatrix} \tag{7}$$

$$G = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ -1 & 0 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -1 \end{bmatrix} \tag{8}$$

Sine data set

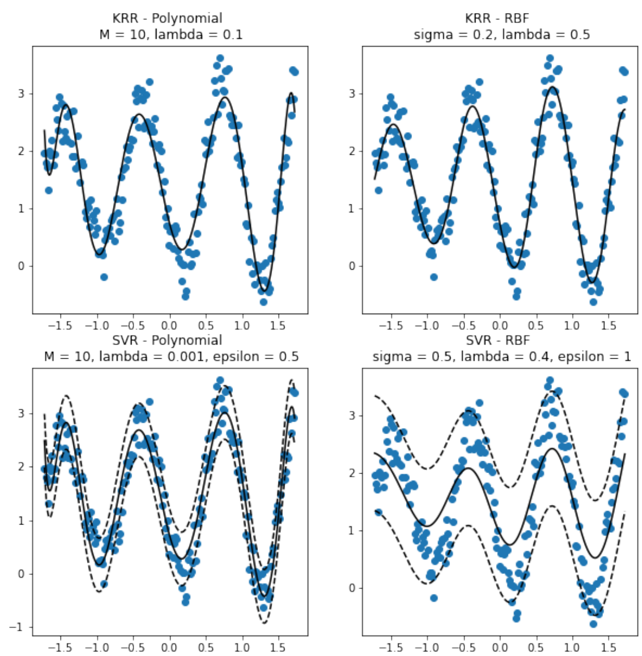

Figure 1. Fittings of SVR and KRR with RBF and Polynomial kernels on the sine data set. The sparse fitting is included in the plots of the SVR model. Blue dots represent the data points from the given data set. The black lines represent the model with a specific kernel. For each combination, the parameter values are reported in the plot titles.

Models Kernelized Ridge Regression and Support Vector Machine have been tested on the sine data set (see Figure 1).

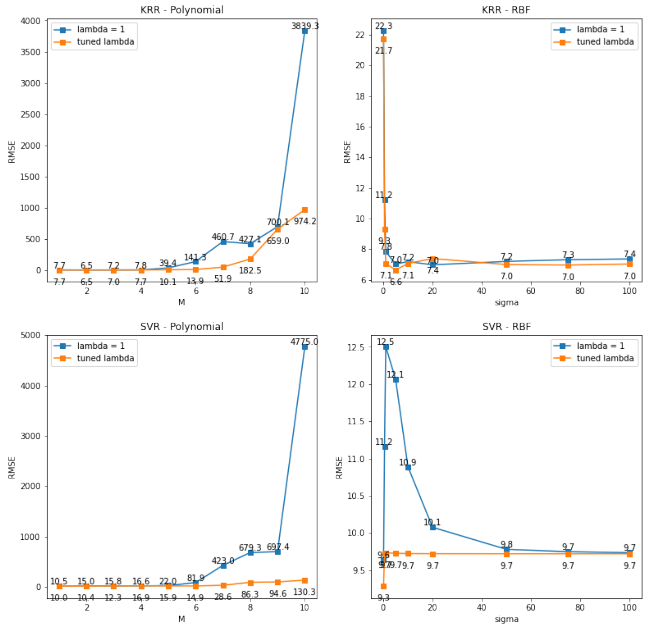

Figure 2. Fittings of SVR and KRR with RBF and Polynomial kernels. On the x-axis are reported the M and sigma parameter values. On the y-axis are reported the RMSEs. Blue plot represents the model with $\lambda = 1$. The orange plot represents the model with the cross-validation tuned λ.

For all four combinations, we were able to find parameters, which create a good fitting of the data.

House data set

Both methods and kernels have been tested on the house data set. On Figure 2 we see the RMSE of different parameter values. On each plot, there are reported RMSEs for $\lambda = 1$ and for cross-validation tuned λ, which was evaluated as the average of 200 lambdas from 200 5-fold cross-validations, where each time the train data was shuffled. Each cross-validation run with different parameters of $\lambda = [0, 0.1, 1, 5, 10, 100, 1000]$. λ tuning was performed using only the train data set. For the SVR model I used $\varepsilon = 0.01$. From 2 we observe that the best performing model was Kernelized Ridge Regression with the RBF kernel. Both methods managed to get low RMSEs. We can observe a difference in the fitting time. KRR is much faster than SVR since the weights are computed using a close-form solution. We also see that the cross-validation tuned λ improved the results of all the methods. In all cases the RMSE of the cross-validation tuned λ was below the RMSE of $\lambda = 1$. In my opinion, the best performing model between the two is the Kernelized Ridge Regression with either of the two kernels.