Machine Learning (ELL409): Assignment 3

Gauri Gupta (2017MT60207)

Support Vector Regression: eps-SVR and RH-SVR

For this assignment, epsilon-SVR and RH- SVR is implemented using the CVXOPT library. The performance of the models is also compared with that of the SVR library used in sklearn.

The dataset is **normalised** before performing any analysis on it. The following are the results obtained on performing **5-fold cross validation** on the given data set.

Hyper-parameters, apart from the kernel hyper-parameters, ε -SVR has two open parameters:

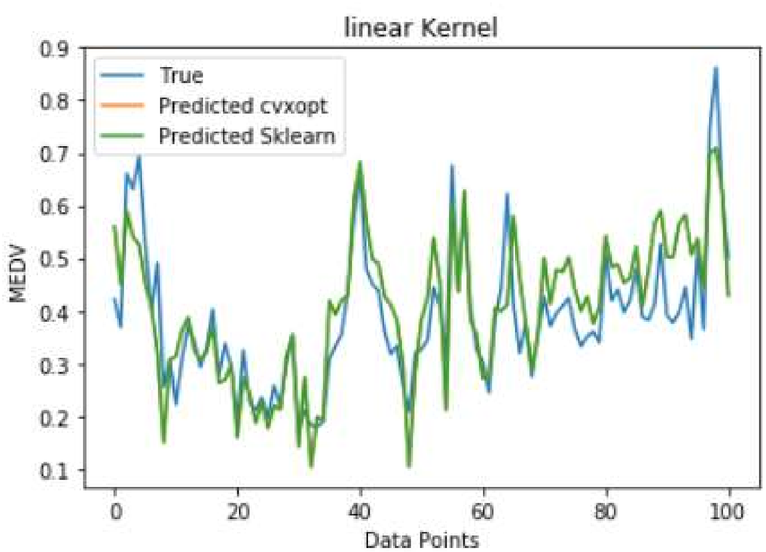
- C: Cost [0 to inf] represents the penalty associated with errors larger than epsilon.

Increasing cost value causes closer fitting to the calibration/training data.

- epsilon represents the minimal required precision

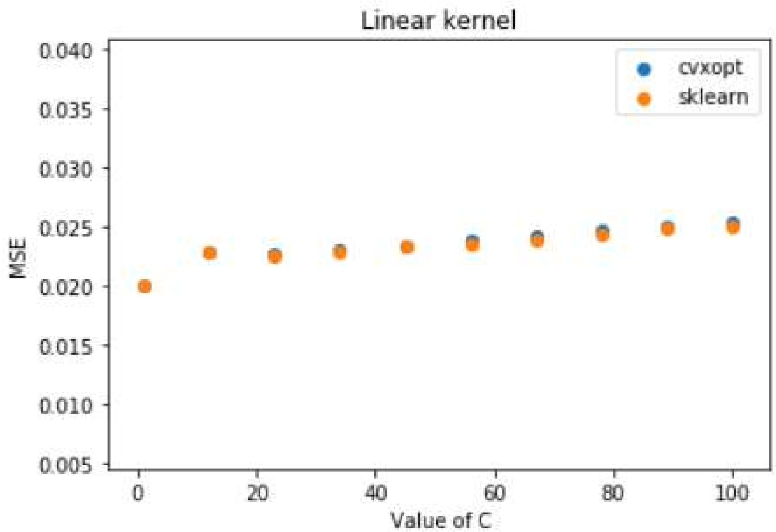
Implementation of Eps – SVR

1. **Linear kernel:**

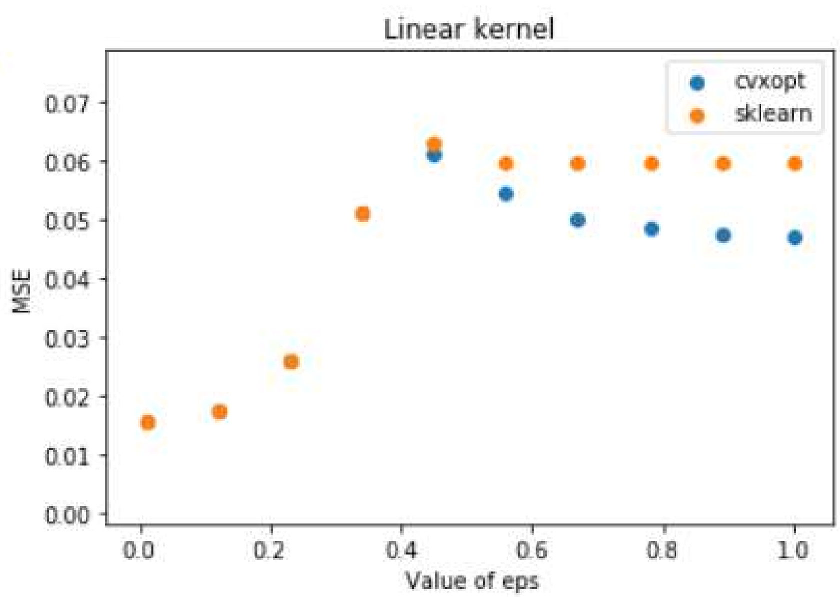


In this graph, the true test values (MEDV) and the ones predicted by the eps-SVR and the sklearn using the linear kernel are plotted. The values predicted by both of these models are almost the same and hence they get overlapped in the graph.

The following graphs were obtained when the hyperparameters C and eps were varied. Mean Square Error (MSE) is plotted against the hyperparamer for both the models (eps-SVR and sklearn)

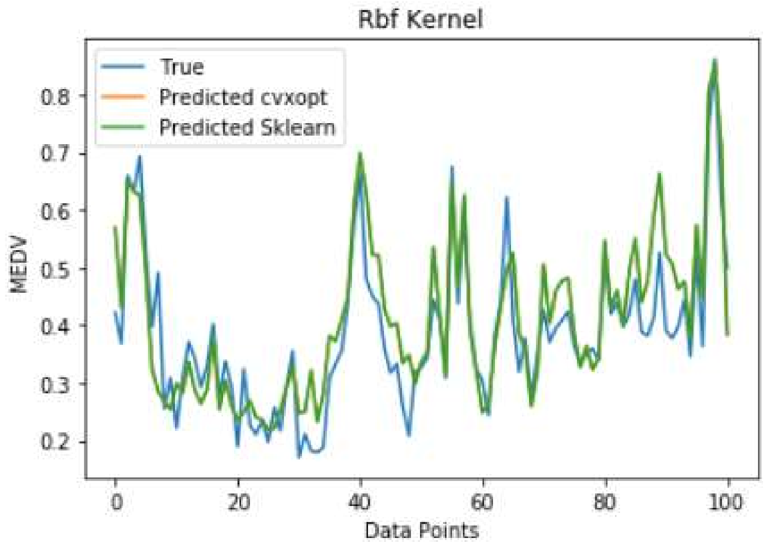


eps = 0.1



C = 1

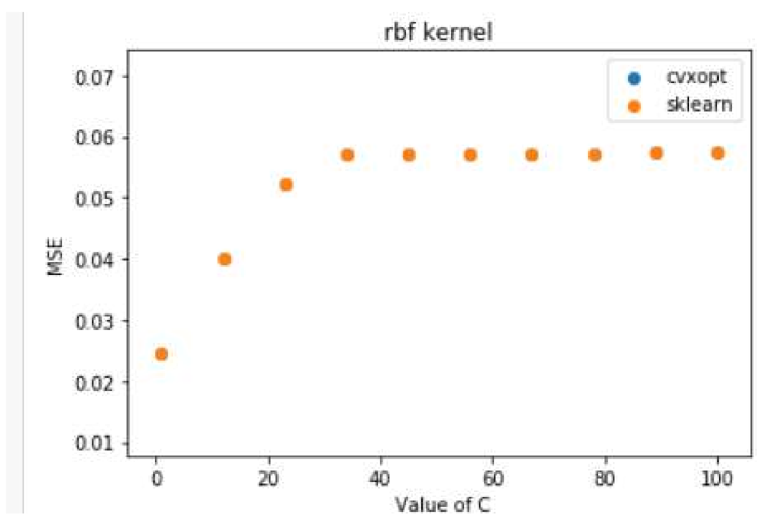
2. RBF kernel:



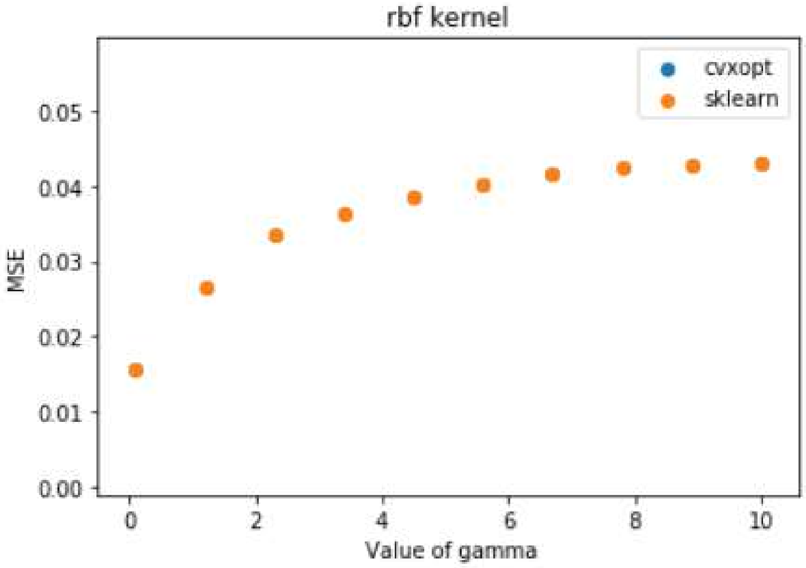
eps = 0.1, gamma = 1

The true test values (MEDV) and the those predicted by the eps-SVR and the sklearn using the rbf kernel are plotted. Tthe values predicted by both of these models are almost the same and hence they get overlapped in the graph.

The following graphs were obtained when the hyperparameters C and gamma were varied. Mean Square Error (MSE) is plotted against the hyperparamer for both the models (eps-SVR and sklearn).



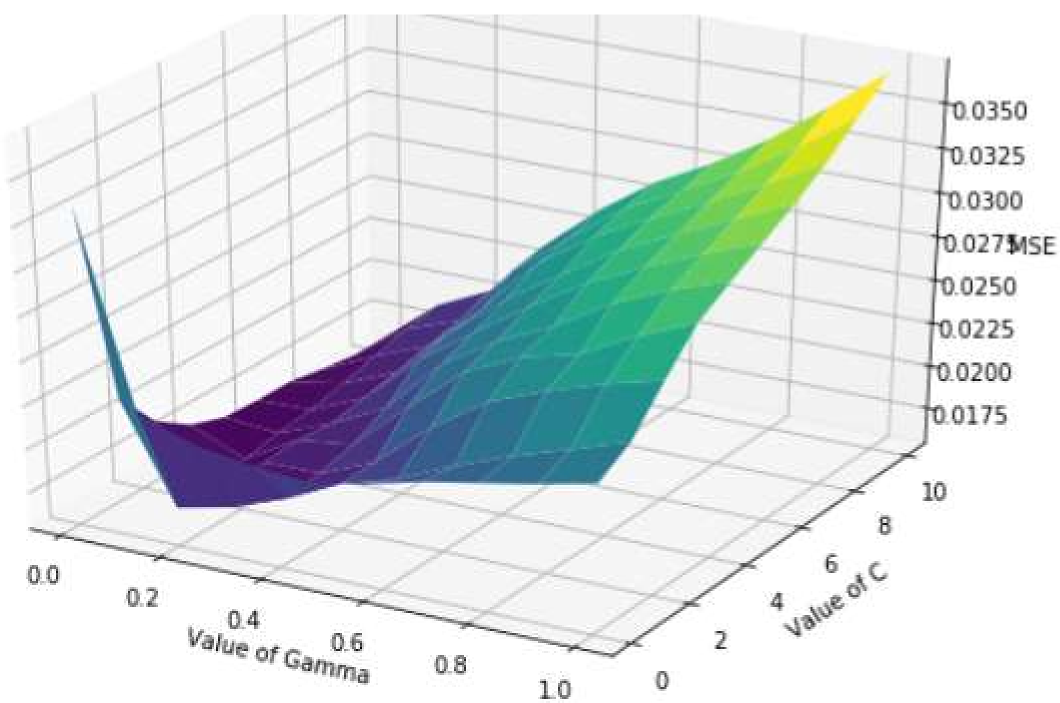
eps = 0.1, gamma = 1

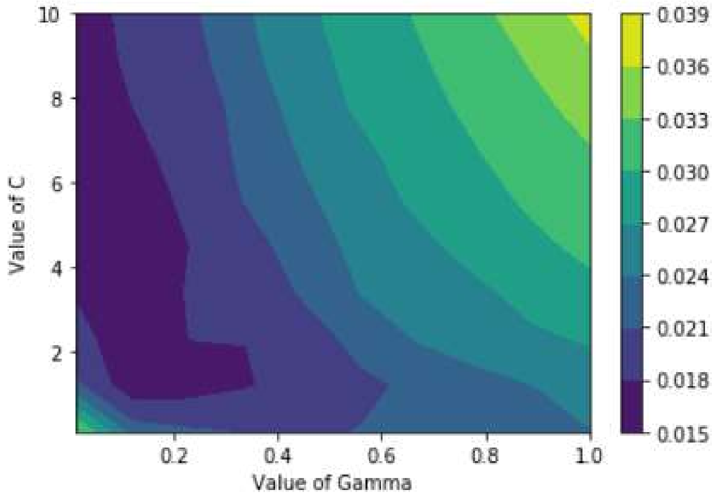


C = 1, eps = 0.1

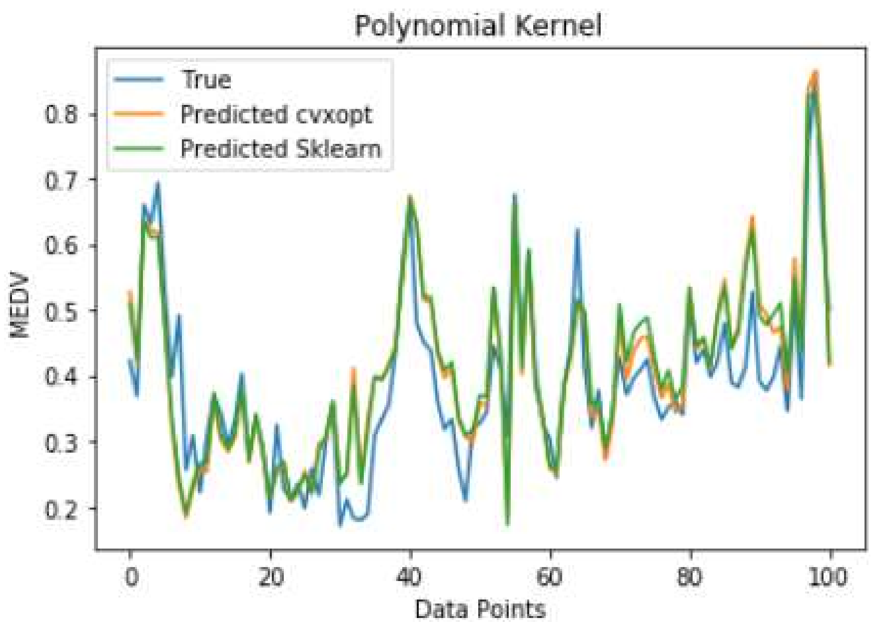
Finding the best parameters for RBF kernel:

```
cvxopt mse :  0.015839397361619597 best_c :  1.2000000000000002 best_gamma :  0.12
sklearn mse :  0.059865955262192884
```



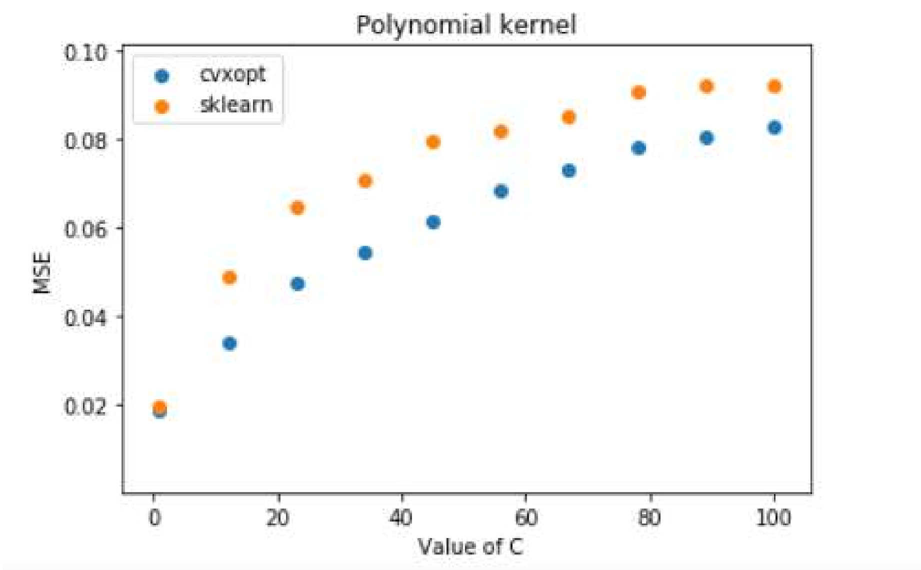


3. Polynomial kernel (degree = 2):

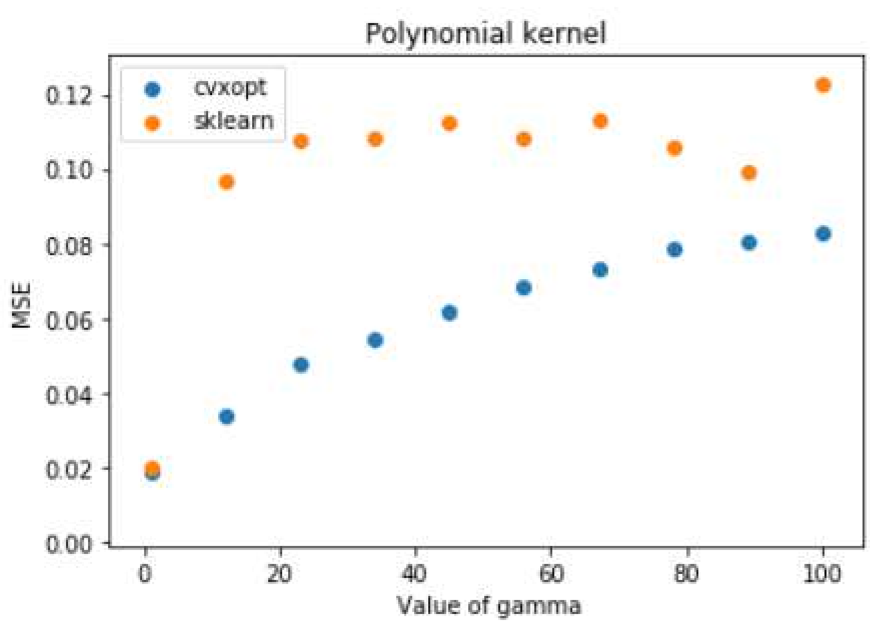


For a given set of training data and test data, this is the plot of the MEDV values predicted by eps-SVR and sklearn vs the true values.

The following graphs were obtained when the hyperparameters C and gamma were varied for degree 2. Mean Square Error (MSE) is plotted against the hyperparamer for both the models (eps-SVR and sklearn)



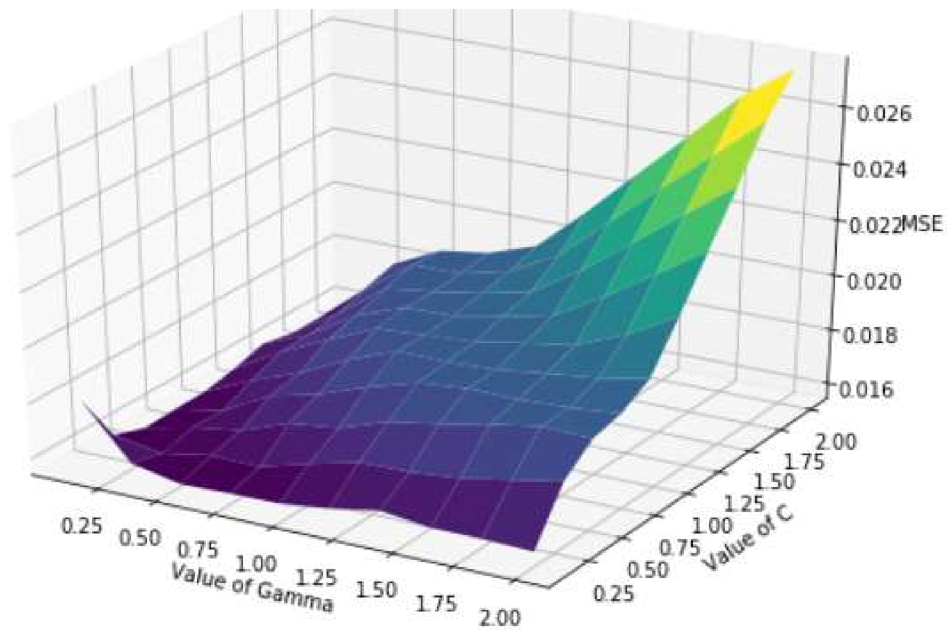
eps = 0.1, gamma =1

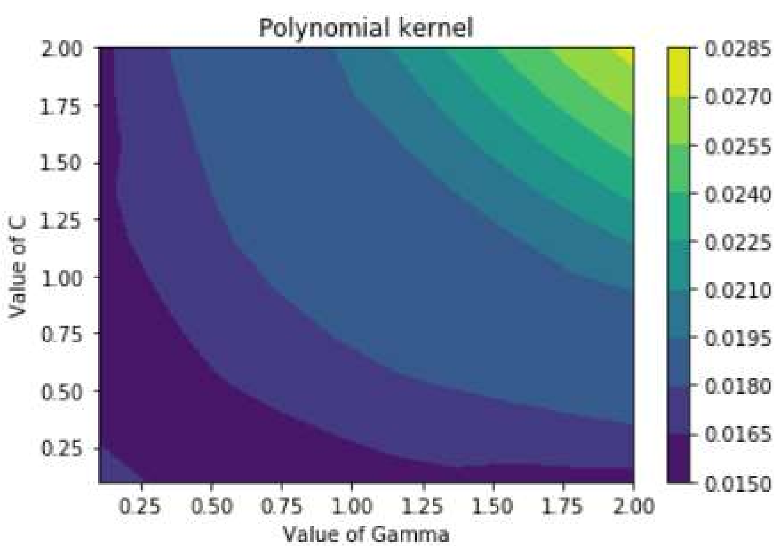


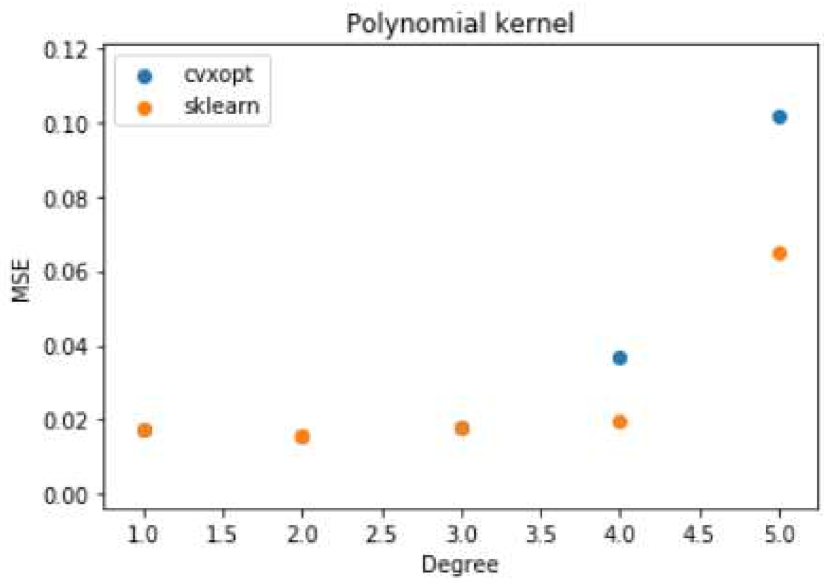
C=1, eps = 0.1

Finding the best parameters:

```
cvxopt mse :  0.015666026908862245 best_c :  0.5222222222222223 best_gamma :  0.1
sklearn mse :  0.059865955262192884
```





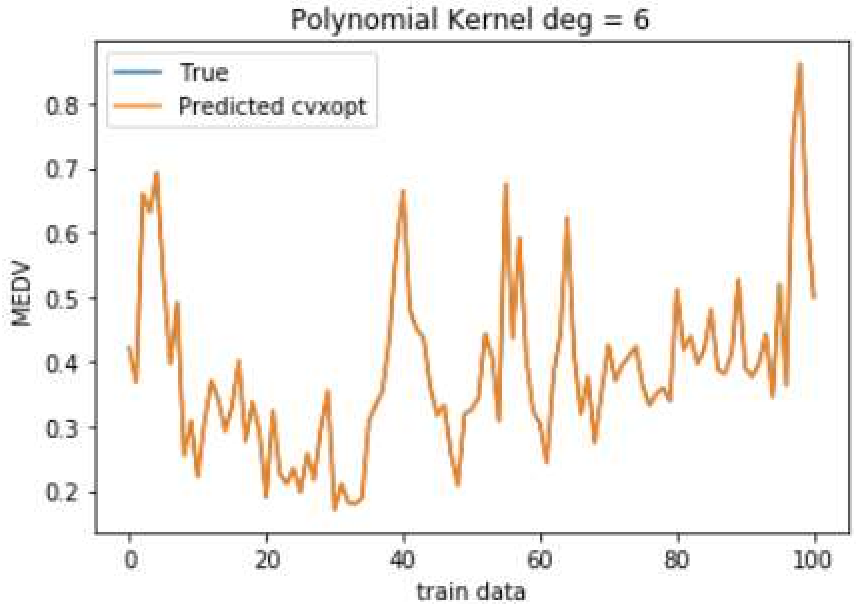


C = 1, eps = 0.1

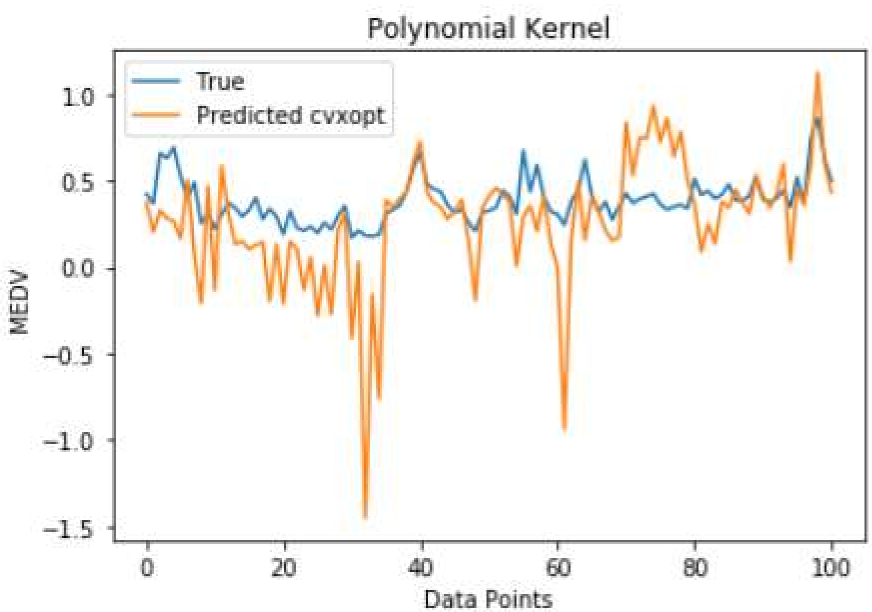
Observations:

- The MSE for both linear and non-linear SVR is less for the eps-SVR using CVXOPT than the standard SVR library.
- The least MSE is obtained while using the polynomial kernel, degree = 2 for the eps-SVR (cvxopt)
- Effect of changing the hyperparameters- On varying the various hyperparameters, such as C, gamma and degree, we could see that as they increased the model complexity increased with due to which model showed overfitting once we went ahead of the optimal points. A higher gamma affects the curvature of the decision boundary due to which overfitting is caused.
- Effect of changing the choice of kernel- We could observe that all the linear, rbf and the polynomial kernel provide almost the same MSE values. This may happen because data itself is linearly separable, all the kernels provide the same accuracy. But we can see that if polynomial kernels with high degree =4 or 5 are used, MSE rises and we tend to overfit the data.

Observing overfitting with polynomial kernel degree = 6



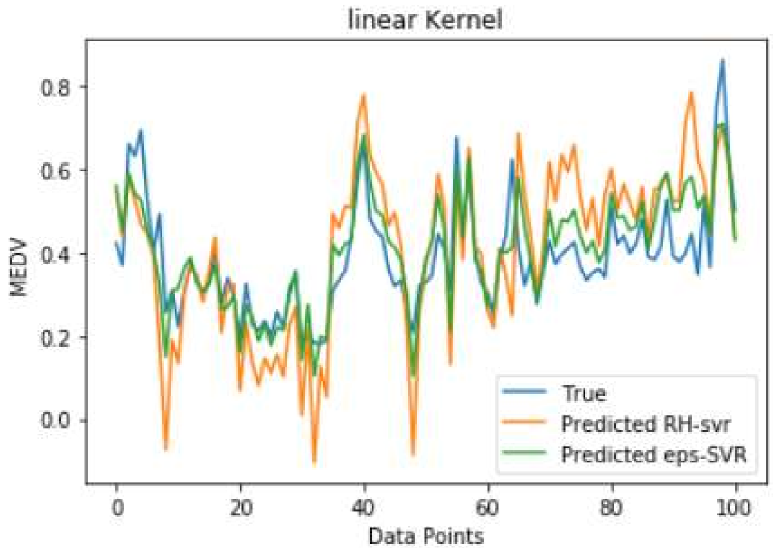
We can see that kernel overfits the training data.



But performs very poorly on the test data.

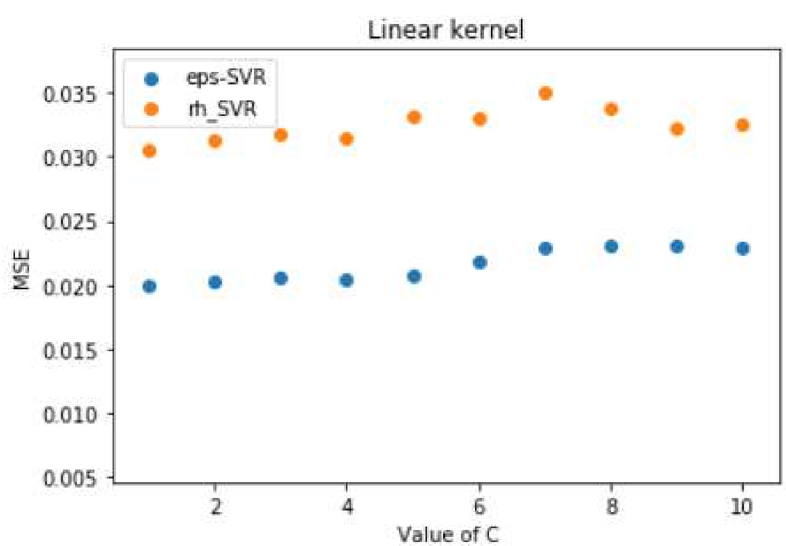
This is in accordance to the observations made earlier. Since the model complexity is increased while using polynomial kernel with degree 6 and hence overfitting.

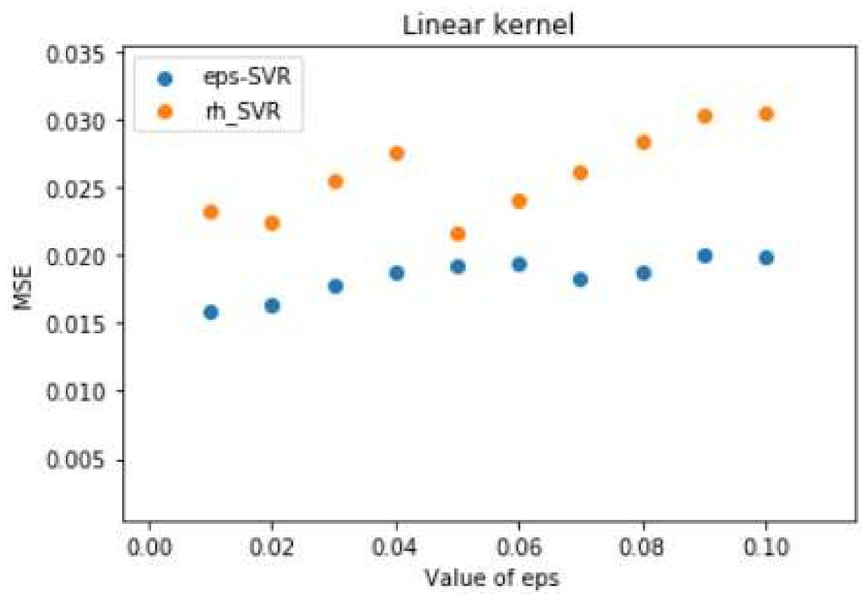
Implementation of RH – SVR



We can see that the values predicted by RH-SVR deviate larger than the eps-SVR from the true values.

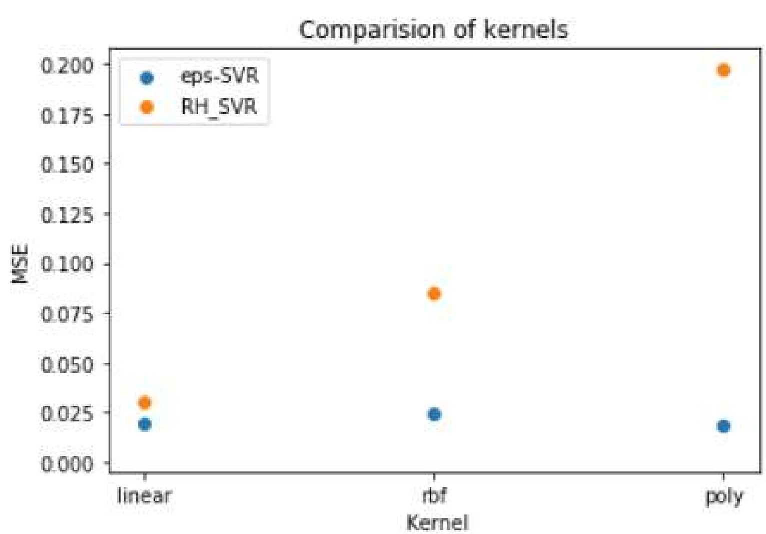
The following graphs were obtained when the hyperparameters C and eps were varied for linear kernel and a comparison is drawn between Rh-SVR and eps-SVR. Mean Square Error (MSE) is plotted against the hyperparameters for both the models (eps-SVR and Rh-svr)





We can see that eps-SVR performs better than rh-SVR for every value of epsilon given.

Comparison between eps-SVR and RH-SVR



C = 1, eps = 0.1, gamma = 1, degree = 2

We can see that eps-SVR performs much better than the RH-SVR for the given data set. Also, both linear and non-linear kernels give the same MSE values for eps-SVR while the MSE increases for rbf and polynomial kernel in the case of RH-kernel.