ELL409

Machine Intelligence and Learning

Assignment – 3 Report

Submitted by:

Karan Mittal(2017ME20671)

Support Vector Regression on Boston House Price Dataset

- Implemented Epsilon Support Vector Regression using general-purpose convex optimization package (CVXOPT) as well as using a customized solver(sklearn). Compared the results obtained using both the methods
- Implemented Reduced Convex Hull Support Vector Regression (RH-SVR) using general-purpose convex optimization package (CVXOPT)

Epsilon Support Vector Regression (without sklearn)

- Used CXOPT package for quadratic programming and implemented it to train the SVR model with various kernels
- Observed the variation in the model performance by changing the parameters of the kernel and hyperparameters (epsilon, C)
- Kernels used:

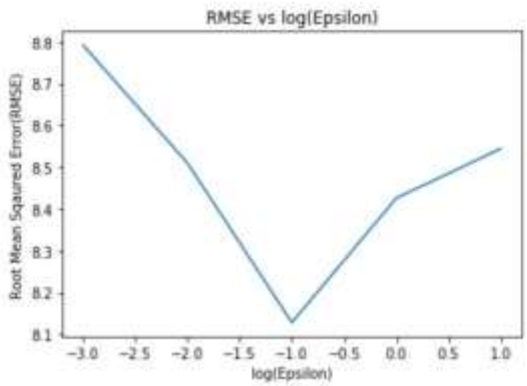
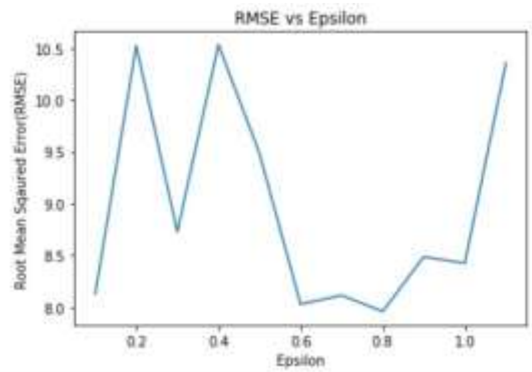
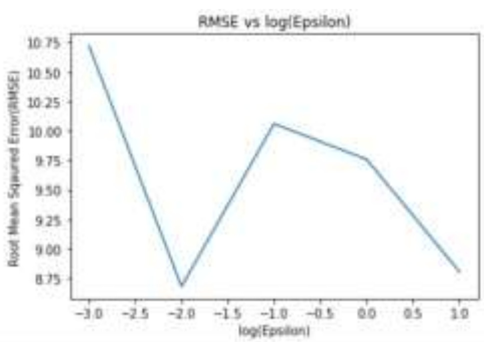
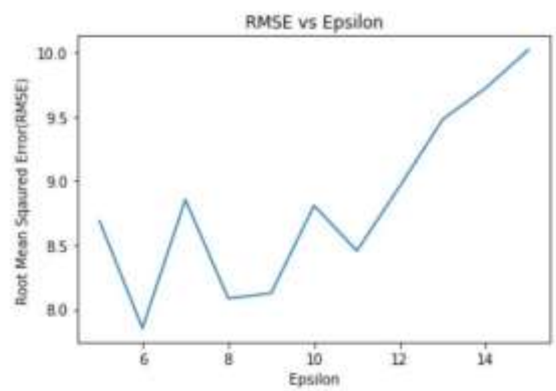
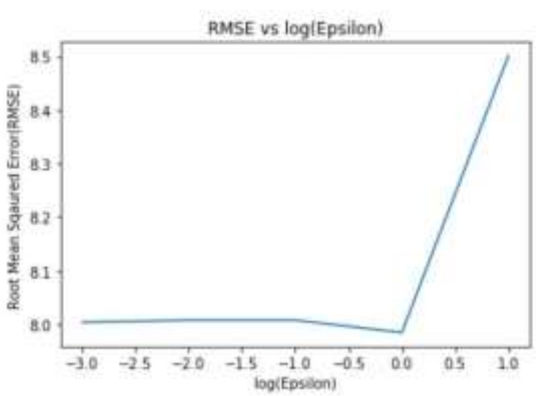
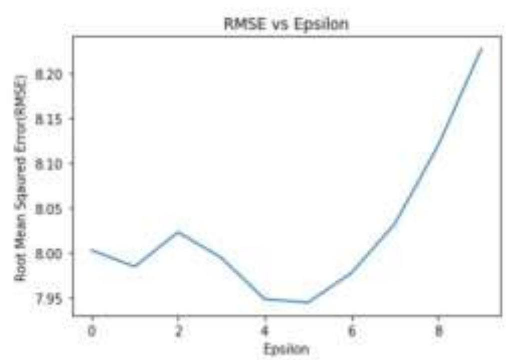
S.No.	Kernel	Formula K(x,y)
1	Linear Kernel	$x^T y$
2	Polynomial Kernel	$(\gamma + x^T y)^d$
3	RBF Kernel	$\exp(- \gamma * \| x - y \|^2)$

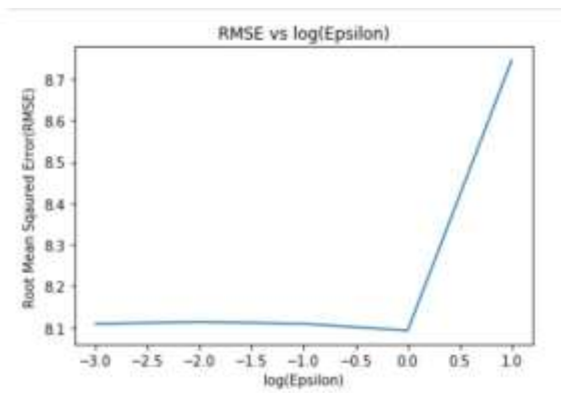
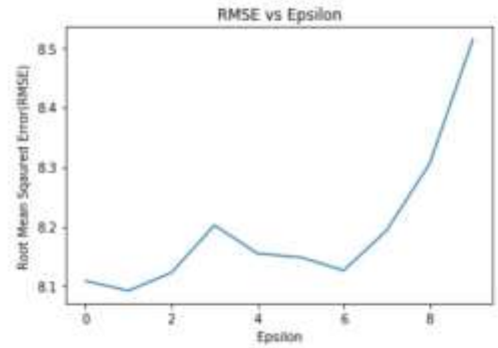
- Metric used to measure the model performance: Root Mean Squared Error
- Used K-Fold Cross Validation to evaluate the performance of the model(K=5)

Variation in RMSE with variation in hyperparameters:

Epsilon

- Plotted RMSE against Epsilon by fixing kernel parameters and taking C=1
- Noted the value of Epsilon which resulted in the minimum value of RMSE

Kernel	RMSE vs log(Epsilon)	RMSE vs Epsilon	Global Minima
Linear **C=1**			**RMSE = 8.1 at Epsilon=0.6**
Polynomial **Gamma=1** **Degree=2** **C=1**			**RMSE = 7.85 at Epsilon=6**
Polynomial **Gamma=1** **Degree=3** **C=1**			**RMSE = 7.94 at Epsilon=5**

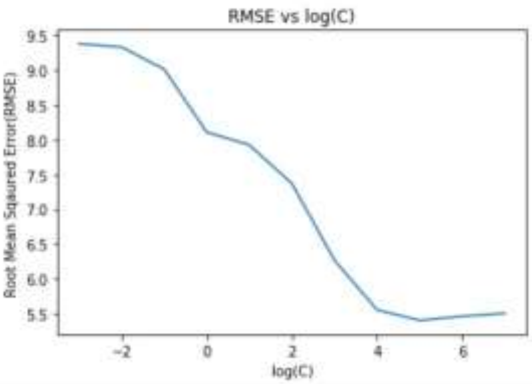
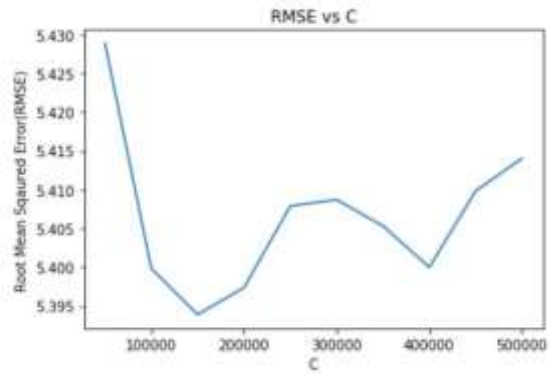
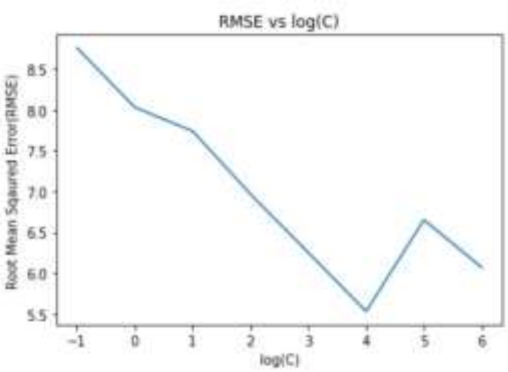
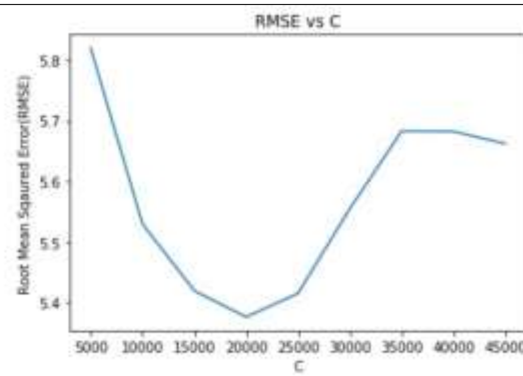
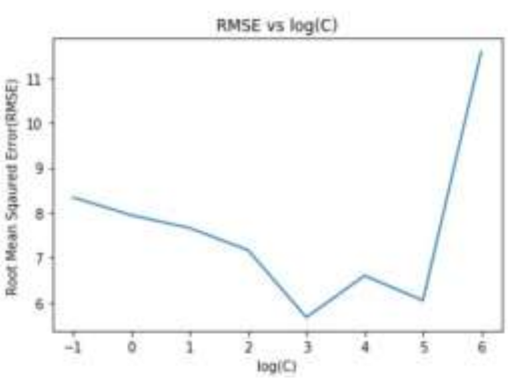
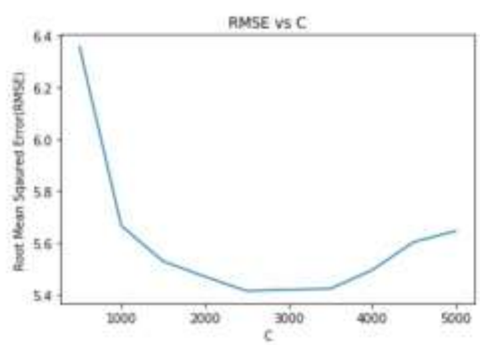
RBF **Gamma=2** **C=1**			**RMSE = 8.09 at Epsilon=1**

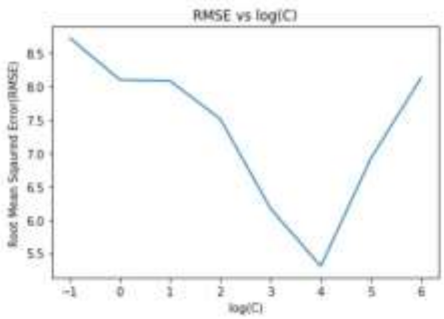
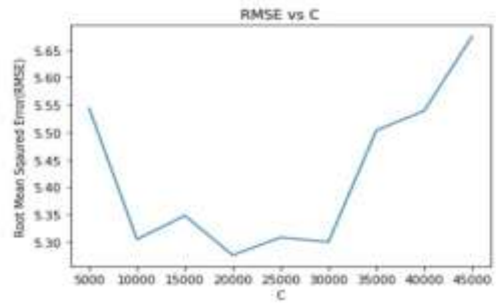
Overfitting and Underfitting Analysis

From the graphs obtained, it can be observed that the RMSE takes very large values for both large values and very small values of epsilon. Minimum value of RMSE is found somewhere between them. Very small value of epsilon means a very thin epsilon tube and large number of support vectors influencing the hyperplane thus causing overfitting. On the other hand, a large value of epsilon means a very wide tube and a very small number of support vectors influencing the hyperplane thus causing underfitting. Therefore, underfitting takes place on the right extreme and overfitting on the left extreme of the graph.

C (Cost of Misclassification of Data)

- Plotted RMSE against C by fixing kernel parameters and taking Epsilon equal to the value obtained in previous analysis
- Noted the value of C which resulted in the minimum value of RMSE

<table>
<tr><th>Kernel</th><th>RMSE vs log(C)</th><th>RMSE vs C</th><th>Global Minima</th></tr>
<tr><td>Linear
Epsilon=0.8</td><td>
</td><td>
</td><td>RMSE = 5.39 at C=150000</td></tr>
<tr><td>Polynomial
Gamma=1
Degree=2
Epsilon=6</td><td>
</td><td>
</td><td>RMSE = 5.37 at C=20000</td></tr>
<tr><td>Polynomial
Gamma=1
Degree=3
Epsilon=5</td><td>
</td><td>
</td><td>RMSE = 5.41 at C=2500</td></tr>
<tr><td></td><td></td><td></td><td></td></tr>
</table>

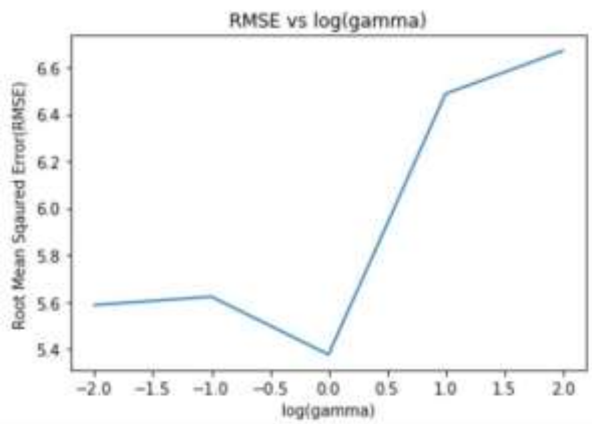
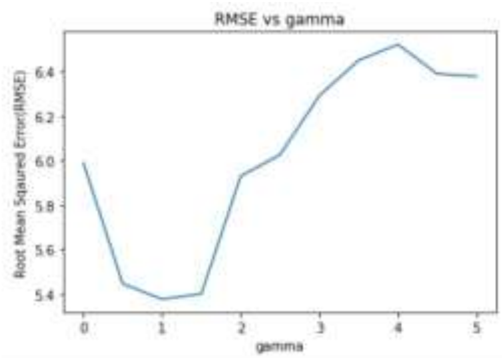
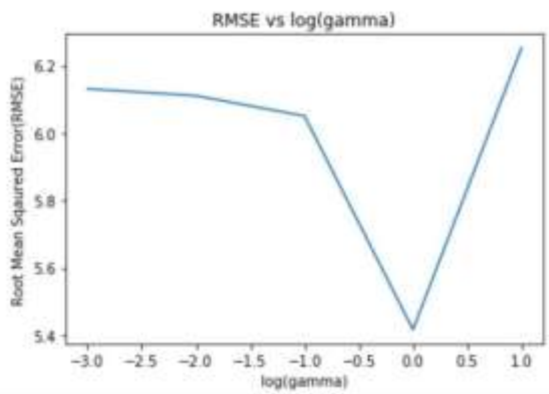
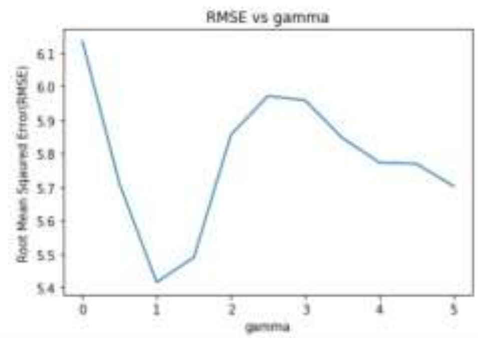
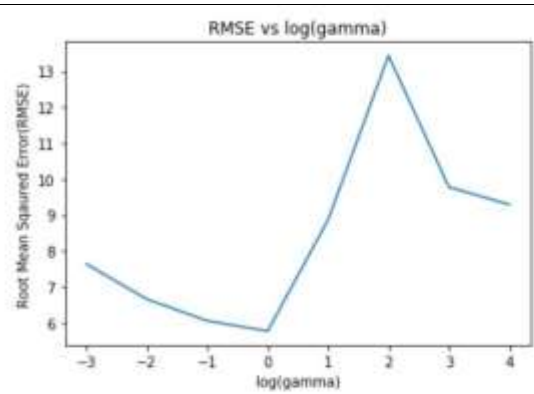
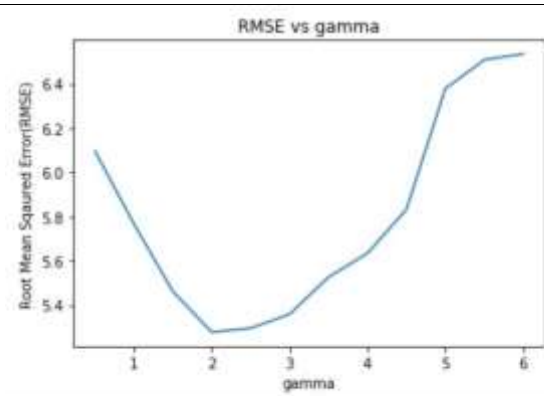
<table>
<tr>
<td>RBF
Gamma=2
Epsilon=1</td>
<td>
</td>
<td>
</td>
<td>RMSE =
5.27 at
C=20000</td>
</tr>
</table>

Overfitting and Underfitting Analysis

C is a trade-off between margin width and misclassifications. From the graphs obtained, it can be observed that the RMSE takes very large values for both very large values and small values of C. For very large values of C, a smaller-margin hyperplane will be obtained if that hyperplane does a better job of getting all the training points classified correctly thus causing overfitting. Conversely, a very small value of C will look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points thus causing underfitting. Therefore, overfitting takes places on the right extreme and underfitting on the left extreme of the graph.

Variation in RMSE with variation in Kernel parameters:

- Plotted RMSE against gamma by taking Epsilon and C equal to the value obtained in the previous sections
- Noted the value of gamma which resulted in the minimum value of RMSE

Kernel	RMSE vs log(gamma)	RMSE vs gamma	Global Minima
Polynomial Degree=2 Epsilon=6 C = 20000			**RMSE = 5.37 at gamma=1**
Polynomial Degree=3 Epsilon=5 C = 2500			**RMSE = 5.41 at gamma=1**
RBF Epsilon = 1 C=20000			**RMSE = 5.27 at gamma = 2**

Overfitting and Underfitting Analysis

From the graphs obtained, it can be observed that the RMSE takes very large values for both very large values and small values of gamma.
If gamma is too large, the radius of the area of influence of the support vectors only includes the support vector itself and no amount of regularization with C will be able to prevent overfitting.
When gamma is very small, the model is too constrained and cannot capture the complexity or “shape” of the data. The region of influence of any selected support vector would include the whole training set.
Therefore, overfitting takes places on the right extreme and underfitting on the left extreme of the graph.

Best Regression Model

The best models for different kernels which I could obtain by analyzing the plots are:

Kernel	**Optimal Hyperparameters**	**RMSE**
Linear	Epsilon=0.8, C=150000	5.39
Polynomial	Epsilon=6, C=20000, gamma=1, degree=2	5.37
RBF	Epsilon=1, C=20000, gamma=2	5.27

RBF kernel performed better than the other two kernels while performing Epsilon Support Vector Regression

Results Obtained Using Scikit-learn library

- Used **GridSearchCV** to find the optimal hyperparameters and Kernel Parameters resulting in the most accurate predictions
- Trained the model using **SVR.fit** and performed regression on test data using **SVR.predict**
- Used **KFold** (n_splits = 5) Cross Validation to compute the Root Mean Squared Error(RMSE)
- Compared these results with the one obtained using CVXOPT package

Optimal hyperparameters obtained using GridSearchCV

Kernel	Optimal Hyperparameters	RMSE obtained
Linear	C=100000, Epsilon=5	5.28
Polynomial	C=100000, Epsilon=4, Gamma=2, Degree=2	5.13
RBF	C=100000, Epsilon=3, Gamma=2.5	5.11

RBF kernel performed better than the other two kernels while performing Epsilon Support Vector Regression

Comparison

Difference between minimum RMSE obtained using the CVXOPT package and the Sci-Kit Library is very small (5.27-5.11=0.16). The reason for this difference is that GridSearchCV is a very efficient way of obtaining the optimal hyperparameters which gives it an edge over the model which I developed using the CVXOPT package.

Therefore, the model developed using CVXOPT can be safely used to perform Support Vector Regression on Boston house-price dataset

Reduced Convex Hull Support Vector Regression (RH-SVR)

- Used CXOPT package for quadratic programming and implemented it to train the SVR model with various kernels
- Observed the variation in the model performance by changing the parameters of the kernel and hyperparameters (epsilon, D)
- Kernels used:

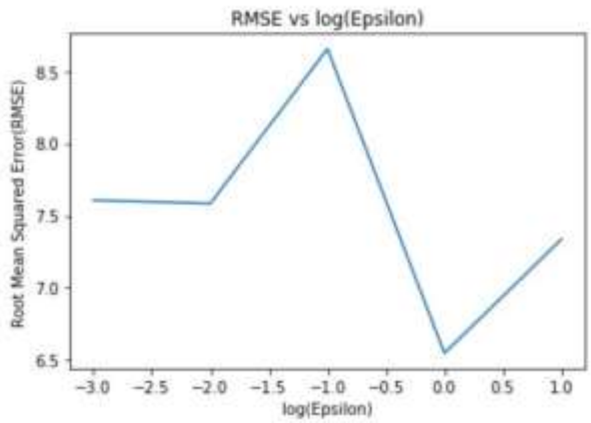
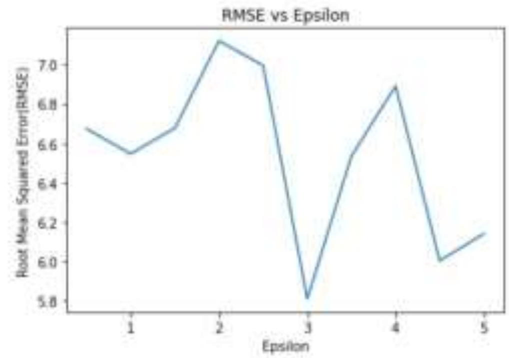
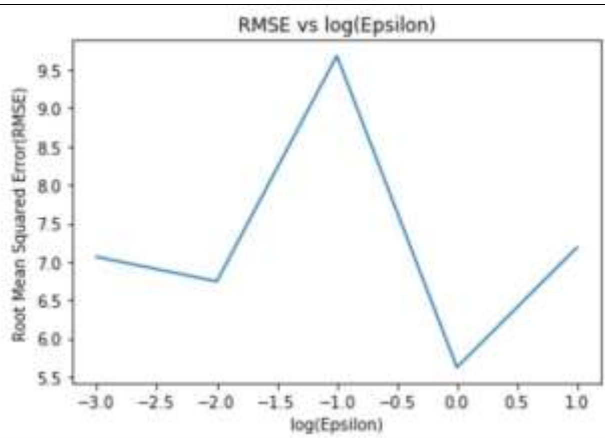
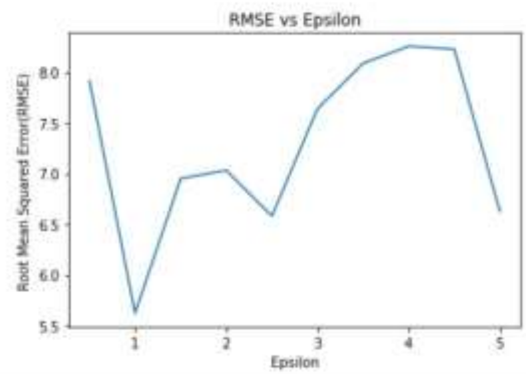
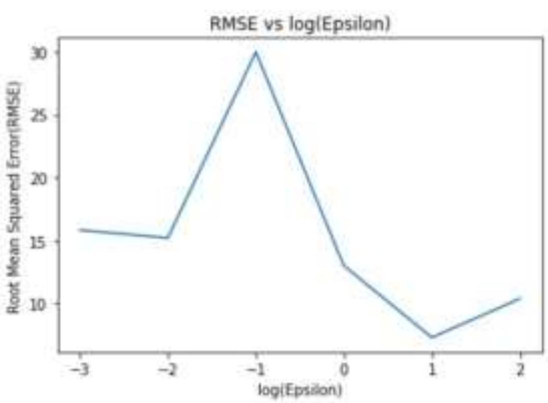
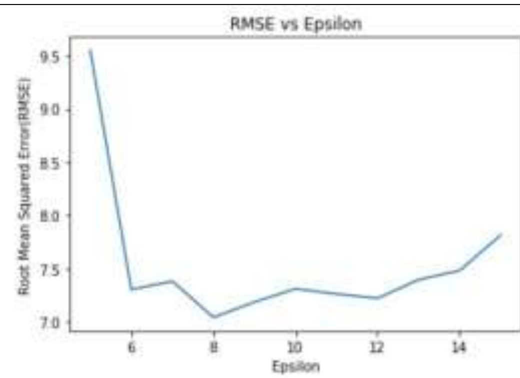
S.No.	Kernel	Formula K(x,y)
1	Linear Kernel	$x^T y$
2	Polynomial Kernel	$(\gamma + x^T y)^d$
3	RBF Kernel	$\exp(-\gamma * \| x - y \|^2)$

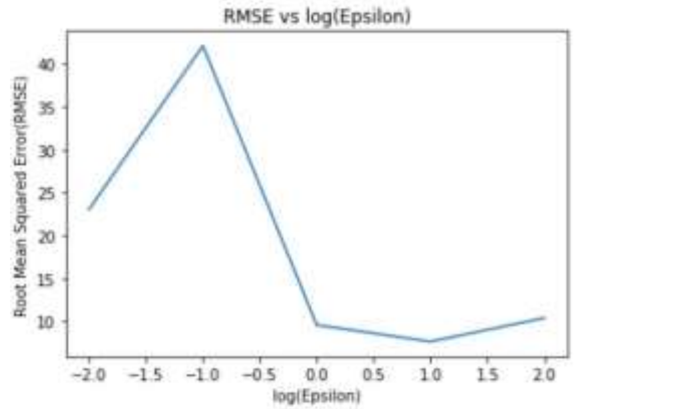
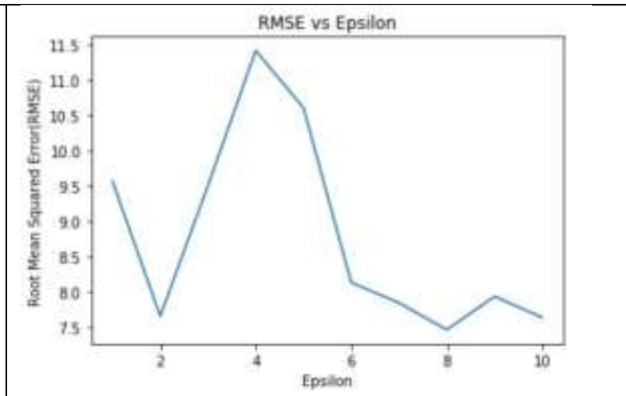
- Metric used to measure the model performance: Root Mean Squared Error
- Used K-Fold Cross Validation to evaluate the performance of the model(K=5)

Variation in RMSE with variation in hyperparameters:

Epsilon

- Plotted RMSE against Epsilon by fixing kernel parameters and taking D=0.5
- Noted the value of Epsilon which resulted in the minimum value of RMSE

Kernel	RMSE vs log(Epsilon)	RMSE vs Epsilon	Global Minima
Linear **D = 0.5**			**RMSE =** **5.80 at** **Epsilon=3**
Polynomial **Gamma=1** **Degree=2** **D=0.5**			**RMSE =** **5.62 at** **Epsilon=1**
Polynomial **Gamma=1** **Degree=3** **D=0.5**			**RMSE =** **7.04 at** **Epsilon=8**

RBF **Gamma=2** **D=0.5**			**RMSE =** **7.46 at** **Epsilon=8**

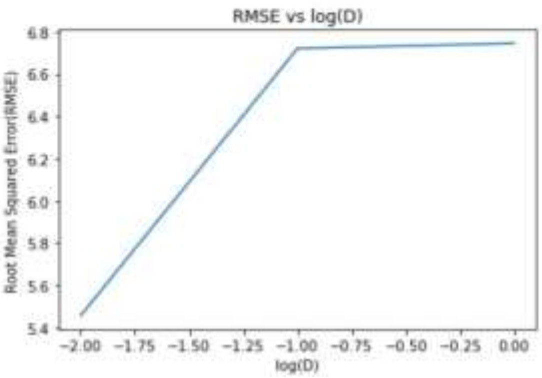
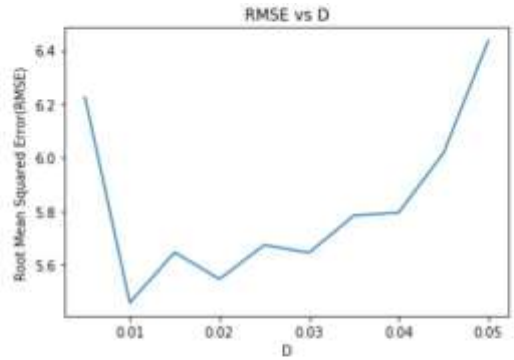
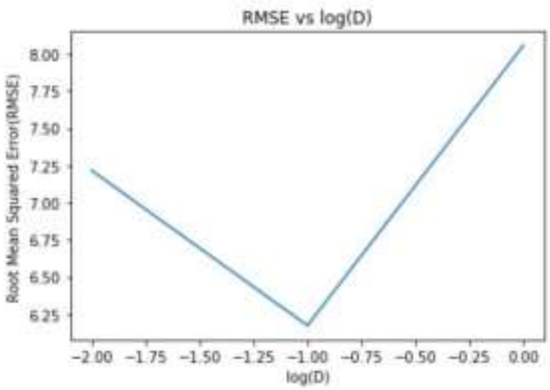
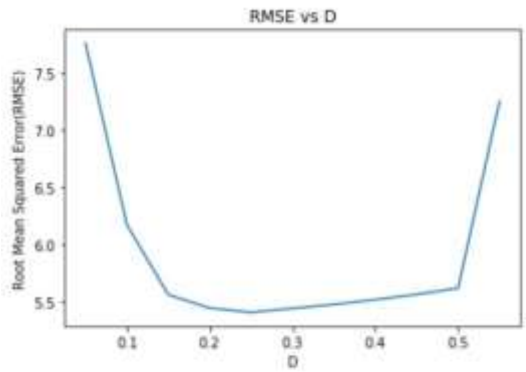
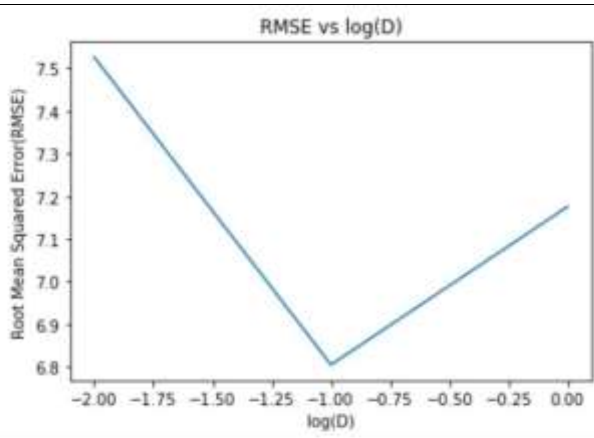
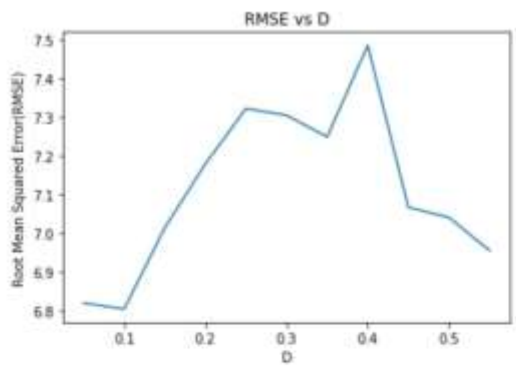
Overfitting and Underfitting Analysis

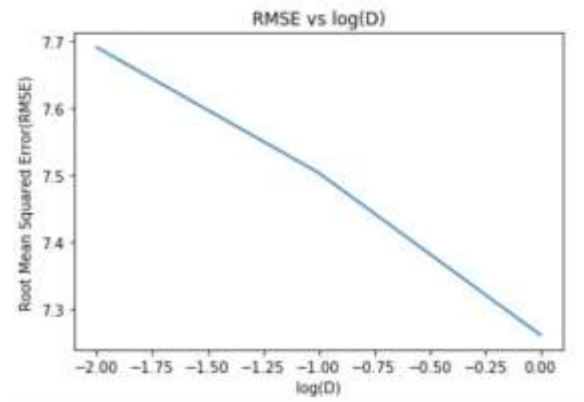
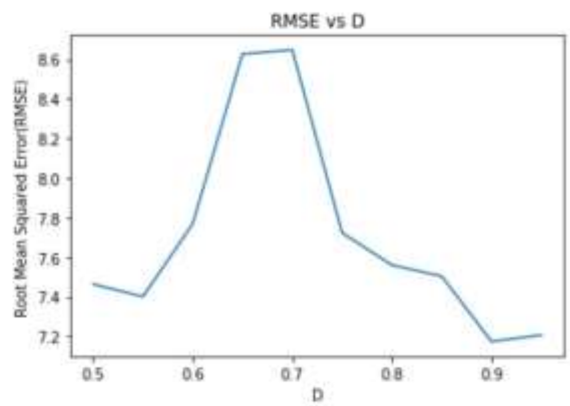
From the graphs obtained, it can be observed that the RMSE takes very large values for both large values and very small values of epsilon. Minimum value of RMSE is found somewhere between them. Very small value of epsilon means a very thin epsilon tube and large number of support vectors influencing the hyperplane thus causing overfitting. On the other hand, a large value of epsilon means a very wide tube and a very small number of support vectors influencing the hyperplane thus causing underfitting. Therefore, underfitting takes place on the right extreme and overfitting on the left extreme of the graph.

D (determines the convex hulls are reduced or not)

D<1: Convex hulls are reduced

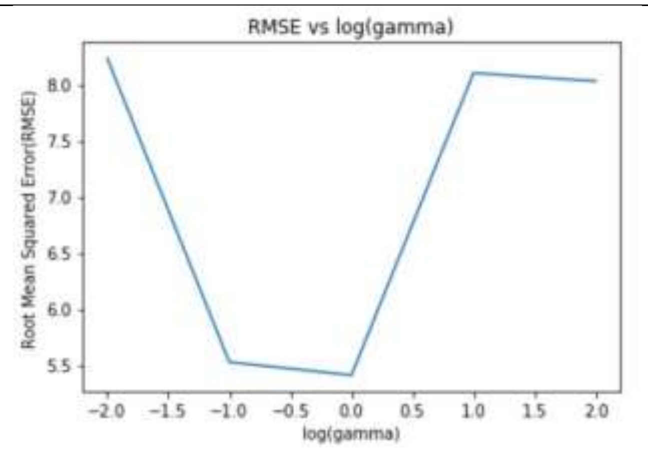
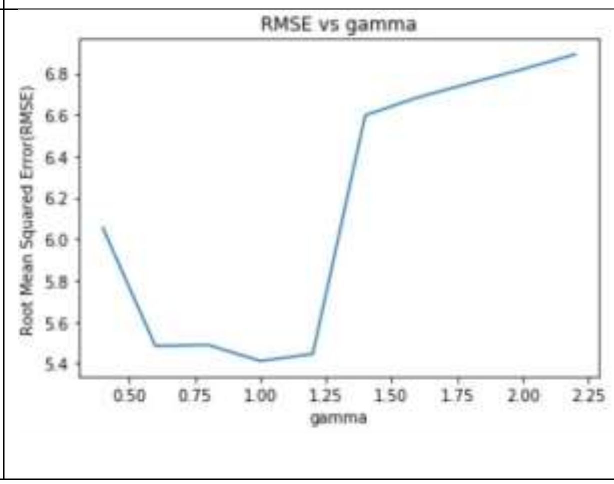
- Plotted RMSE against D by fixing kernel parameters and taking Epsilon equal to the value obtained in previous analysis
- Noted the value of D which resulted in the minimum value of RMSE

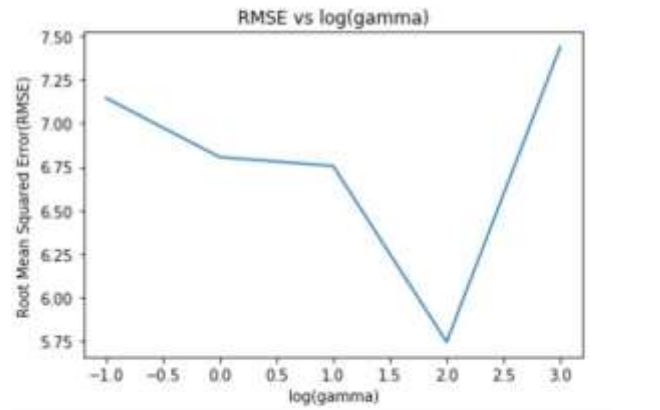
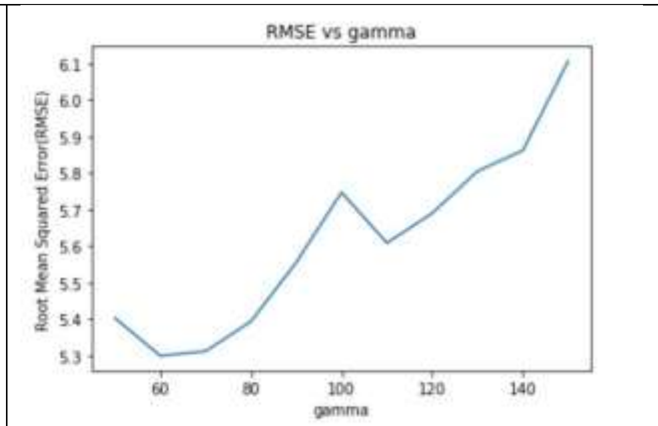
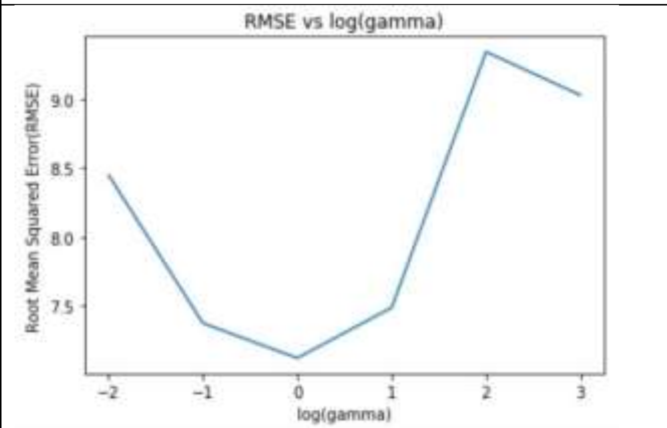
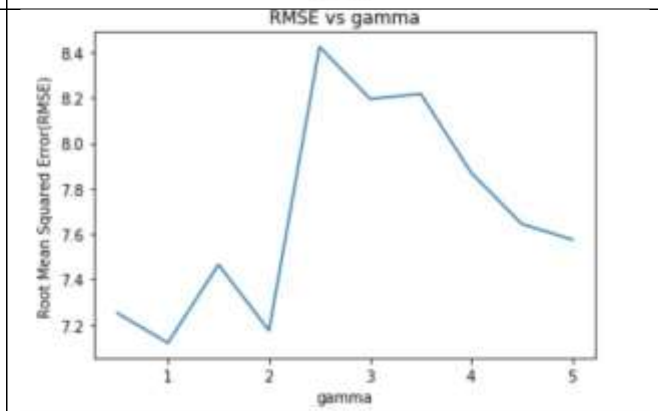
Kernel	**RMSE vs log(D)**	**RMSE vs D**	**Global Minima**
Linear **Epsilon=3**			**RMSE = 5.45 at D=0.01**
Polynomial **Gamma=1** **Degree=2** **Epsilon=1**			**RMSE = 5.41 at D=0.25**
Polynomial **Gamma=1** **Degree=3** **Epsilon=8**			**RMSE = 7.04 at D=0.1**

RBF **Gamma=2** **Epsilon=8**			**RMSE =** **7.17 at** **D=0.9**

Variation in RMSE with variation in Kernel parameters:

- Plotted RMSE against gamma by taking Epsilon and D equal to the value obtained in the previous sections
- Noted the value of gamma which resulted in the minimum value of RMSE

Kernel	RMSE vs log(gamma)	RMSE vs gamma	Global Minima
Polynomial **Degree=2** **Epsilon=1** **D=0.25**			**RMSE =** **5.41 at** **gamma=1**

Polynomial Degree=3 Epsilon=8 D=0.1			**RMSE = 7.04 at gamma=60**
RBF Epsilon=8 D=0.9			**RMSE = 7.11 at gamma=1**

Overfitting and Underfitting Analysis

From the graphs obtained, it can be observed that the RMSE takes very large values for both very large values and small values of gamma.
If gamma is too large, the radius of the area of influence of the support vectors only includes the support vector itself and no amount of regularization with C will be able to prevent overfitting.
When gamma is very small, the model is too constrained and cannot capture the complexity or “shape” of the data. The region of influence

of any selected support vector would include the whole training set. Therefore, overfitting takes places on the right extreme and underfitting on the left extreme of the graph.

Best Regression Model

The best models for different kernels which I could obtain by analyzing the plots are:

Kernel	**Optimal Hyperparameters**	**RMSE**
Linear	Epsilon=3, D=0.01	5.45
Polynomial	Epsilon=1, D=0.25, gamma=1, degree=2	5.41
RBF	Epsilon=8, D=0.9, gamma=1	7.11

Polynomial kernel performed better than the other two kernels while performing Reduced Convex Hull Support Vector Regression (RH-SVR)

References

- Bi, J. and Bennett, K.P. 2003. A geometric approach to support vector regression. Neurocomputing 55:79-108
- Smola, A.J. and Scholkopf, B. 2003. A Tutorial on Support Vector Regression. Statistics and Computing 14: 199–222