HOMEWORK2

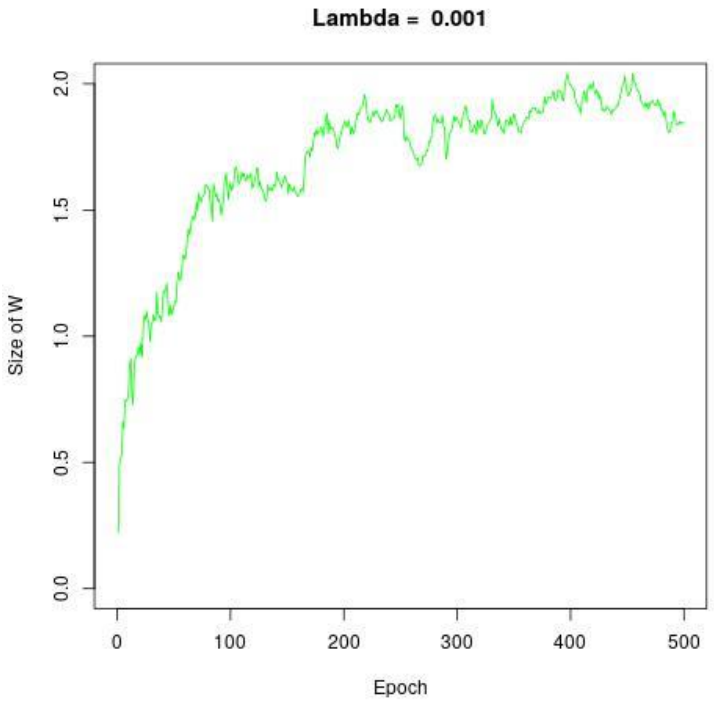
Following assignment asked us to implement support vector machine using Stochastic gradient descent. The dataset has been firstly pre-processed by removing the records having missing values, extracting only the continuous features, encoding the label to numerical form.

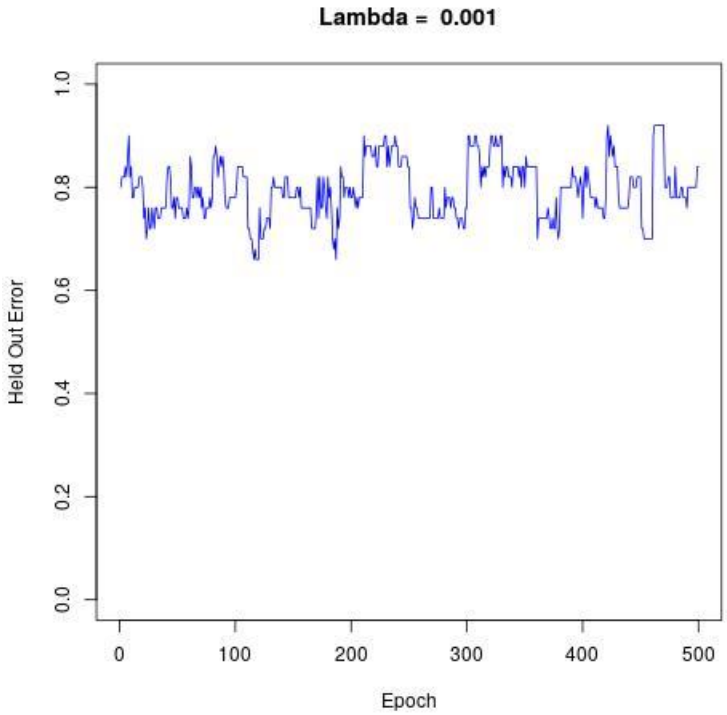
The dataset was splitted into 80-10-10 ratio where 80% was used for training, 10% was used for validation and rest 10% was used for testing.

For Stochastic gradient descent we have used 50 epochs, each having 300 steps and a season of 30 steps after which the accuracy of the model and the magnitude of coefficient was recorded which was used for plotting. The whole process was repeated on different value of regularization constant and based on the accuracy on validation data we picked the lambda which gave the maximum accuracy. This lambda was used to get the generalization error on test data.

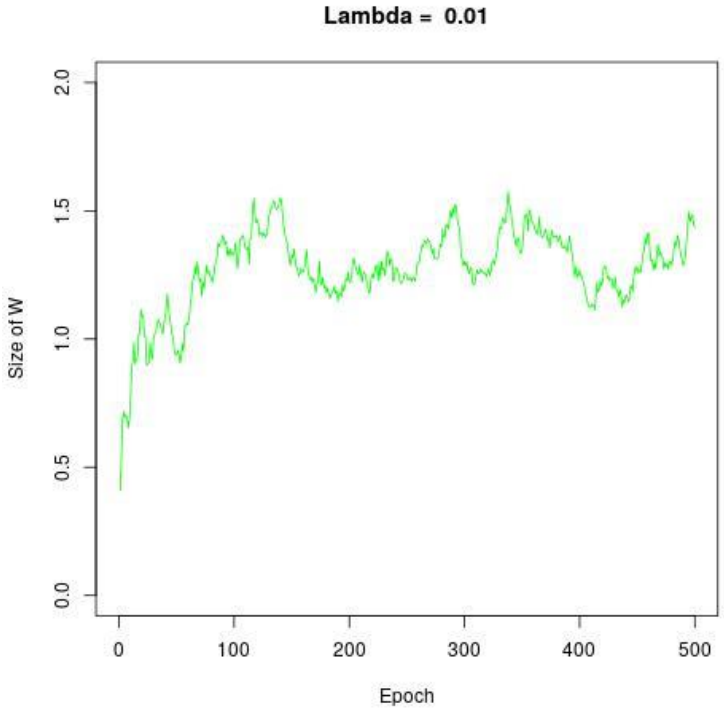
->Following are the outputs of the error on validation data and magnitude of coefficients for each value of lambda:

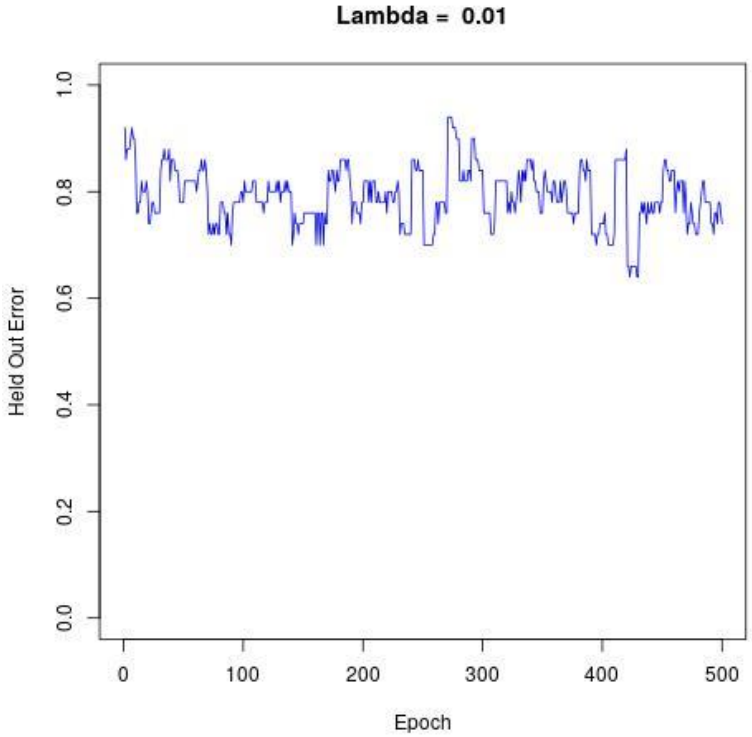
For lambda= 0.001:



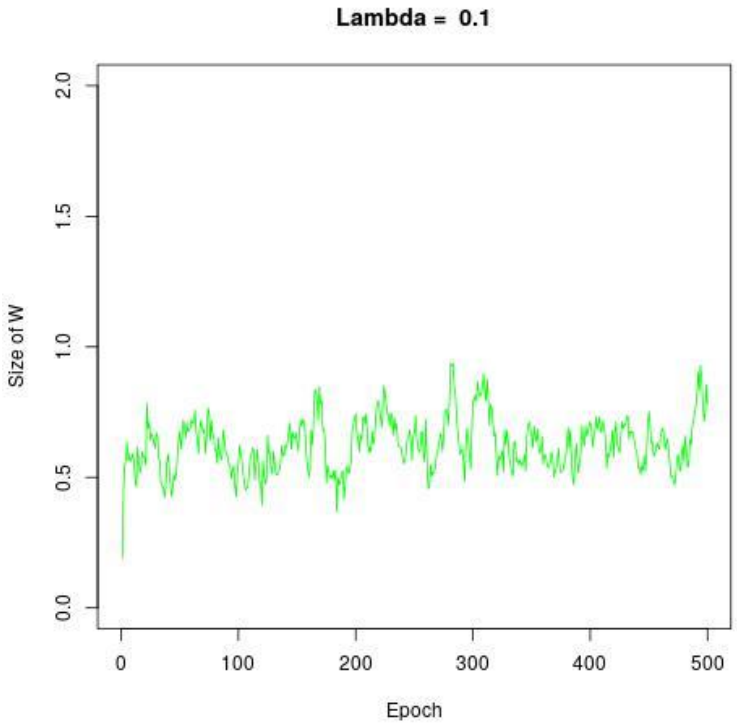


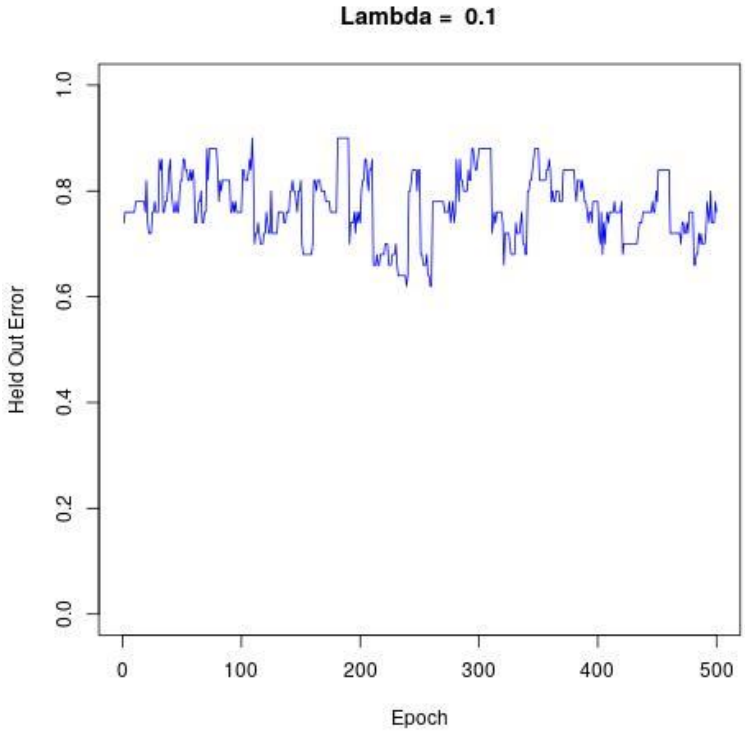
For Lambda=0.01



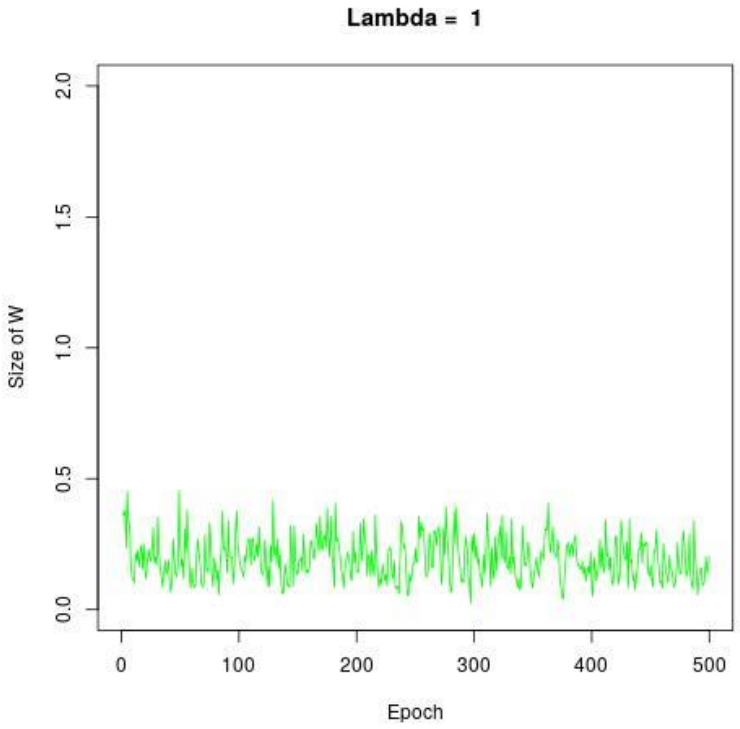


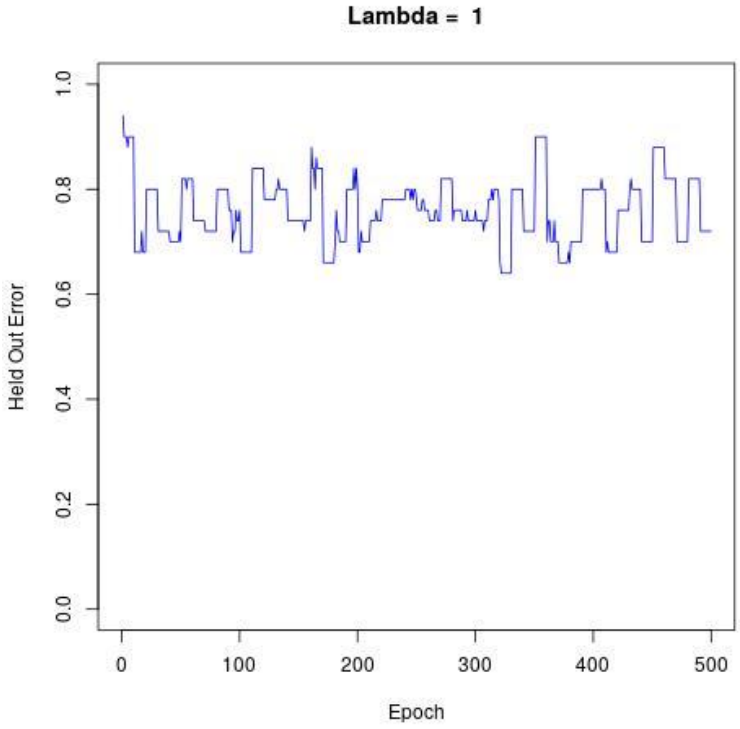
For Lambda =0.1





For Lambda =1


Lambda = 1
Size of W
0.0
0.5
1.0
1.5
2.0
0
100
200
300
400
500
Epoch


Lambda = 1
Held Out Error
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400
500
Epoch

->By running the algorithm, we achieved maximum accuracy with lambda =0.001, so our best estimate of **regularization constant is 0.001.** The accuracy score on lambda=0.01 was close to this score but as we kept on increasing the value of lambda it tends to penalize the features more due to which the contribution of features started to decrease causing under fitting which negatively impacted the generalization error. On the other hand, having a relatively low value of lambda didn't really penalize the features due to which overfitting started occurring.

->Test accuracy reported on 10% test data= 81.82%