

COVID - 19

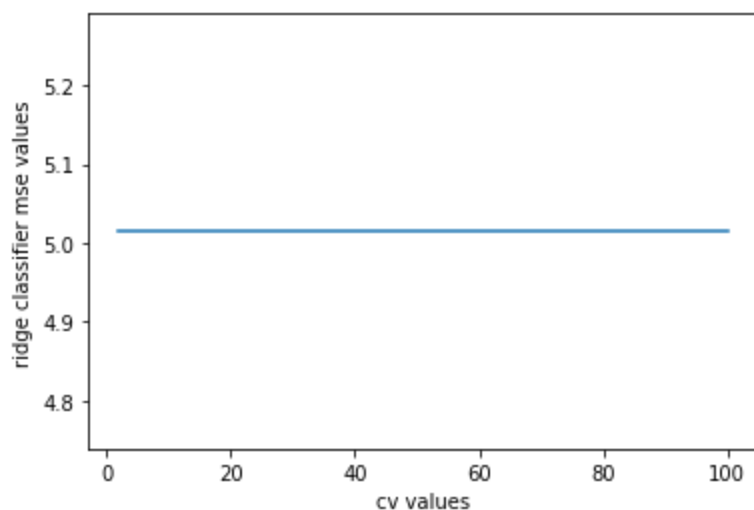
1. MOL2VEC:

Performance of Classifiers with different hyperparameters:

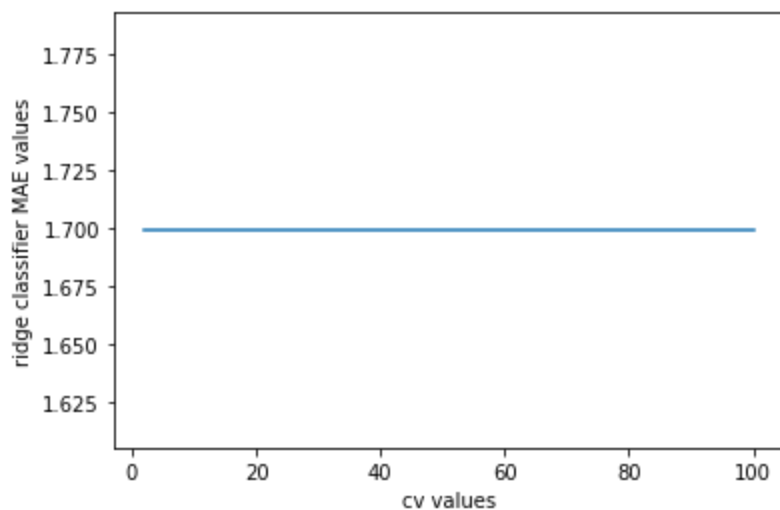
RIDGECV CLASSIFIER:

Ridge regression with built-in cross-validation.

Different values of cv: MSE values



Different values of CV : MAE values

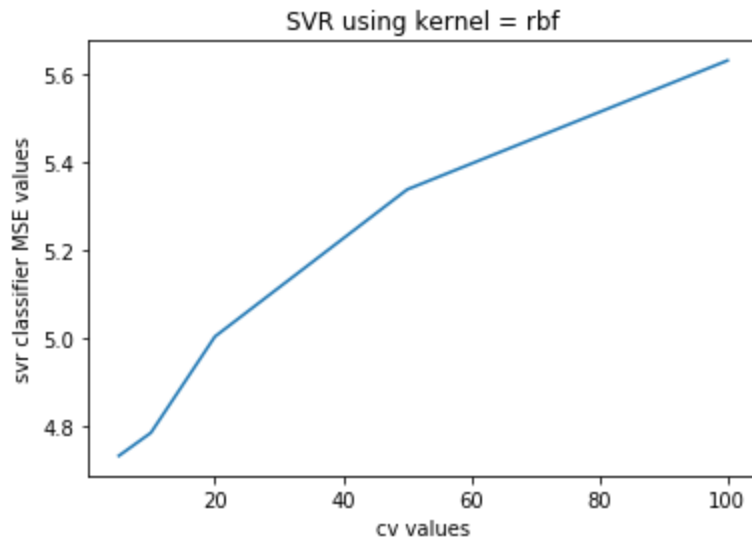


Observation:

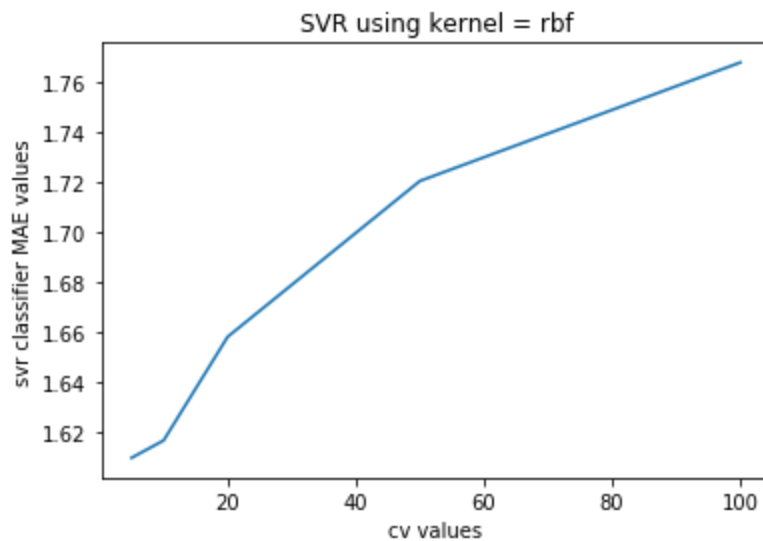
As we can clearly see that for all cross-validation splits, the MSE and MAE values are the same, hence we can use any value for the 'CV' parameter.

SVR CLASSIFIER

Different values of C and kernel = "rbf" : MSE values



Different values of C and kernel = "rbf" : MAE values

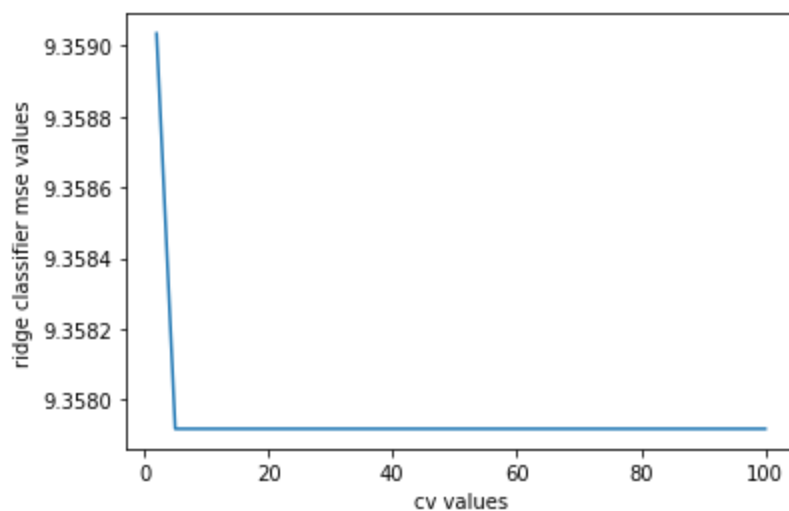


2. No of atoms such as C, O, N, Cl and other heavy atoms

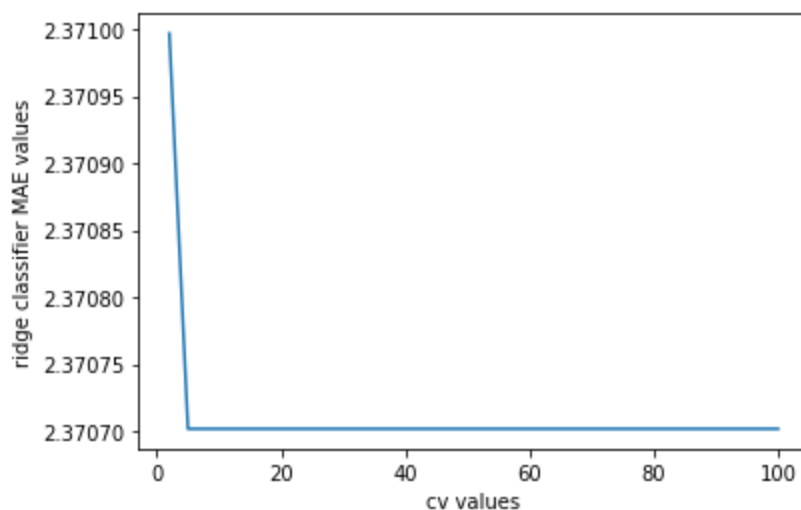
Performance of Classifiers with different hyperparameters:

RIDGE CV CLASSIFIER:

Different values of cv: MSE values



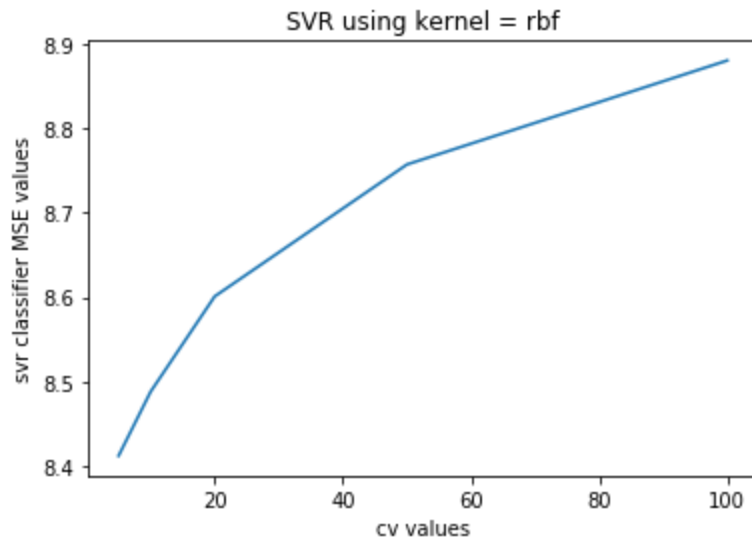
Different values of CV : MAE values



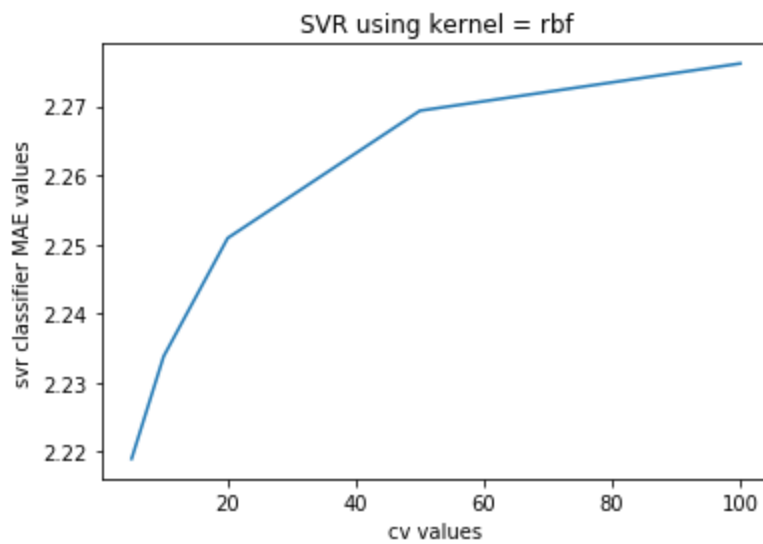
Observation: After $cv = 4$, the value of MSE and MAE is constant. Hence i selected the value of $cv = 10$.

SVR CLASSIFIER

Different values of C and kernel = "rbf" : MSE values



Different values of C and kernel = "rbf" : MAE values

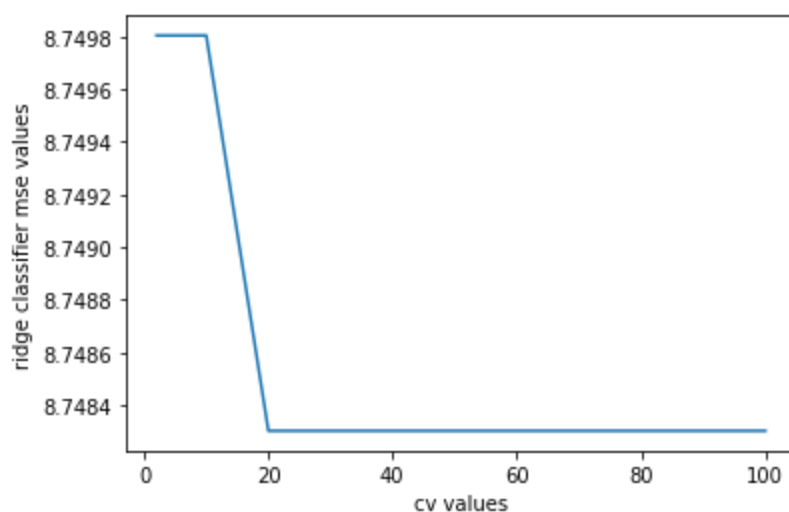


3. Including some more features in the previous dataset such as Molecular wt, No of valence electrons and No of heterogeneous atoms

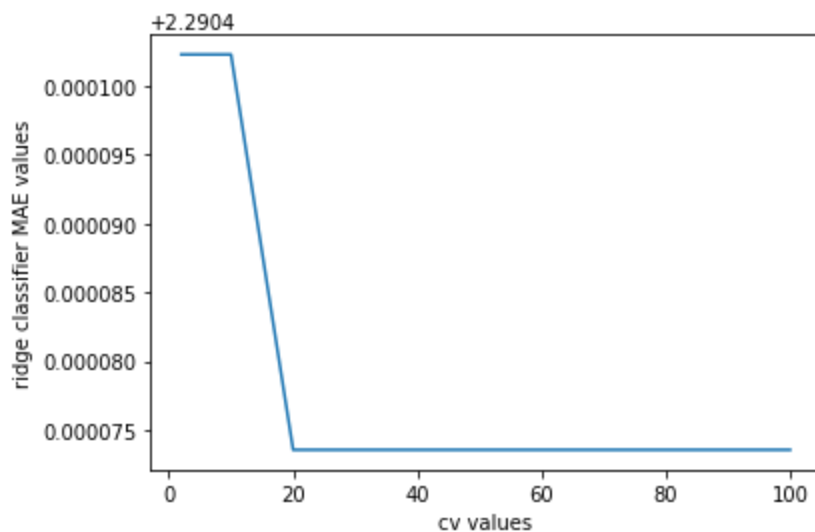
Performance of Classifiers with different hyperparameters:

RIDGE CV CLASSIFIER:

Different values of cv: MSE values



Different values of CV : MAE values

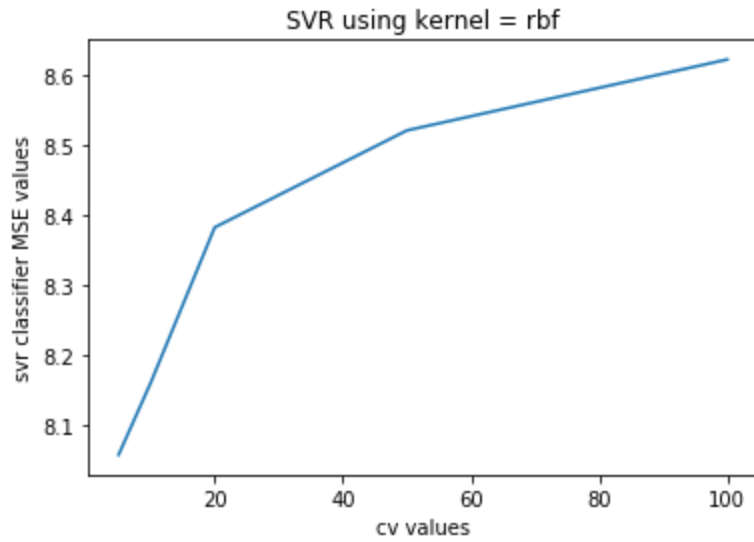


Observation:

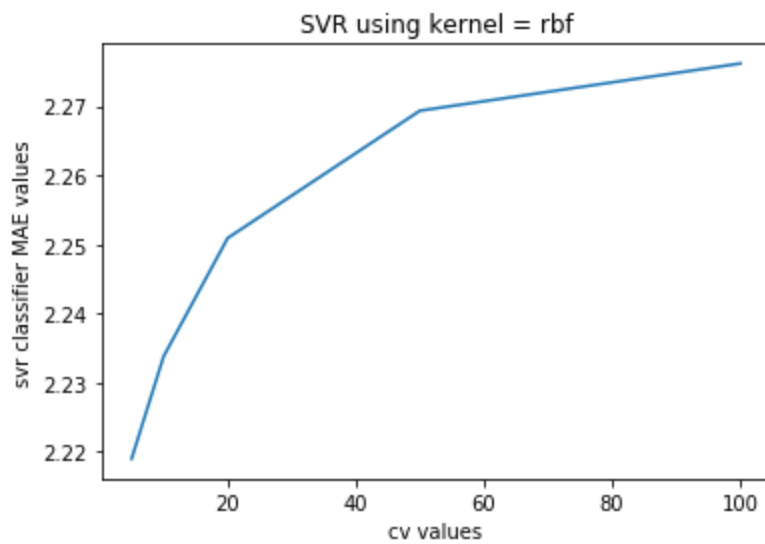
Here the value of MSE and MAE is constant after CV = 20.

SVR CLASSIFIER

Different values of C and kernel = "rbf" : MSE values



Different values of C and kernel = "rbf" : MAE values



Observation:

SVR performs much better than RidgeCV and with increasing values of C, the MAE and MSE values also increase. The best MSE and MAE values are obtained at value 100. Including 'epsilon = 1' in the hyperparameter outperformed every other classifier and svr itself without epsilon being used. Hence I used hyperparameter C = 100 and epsilon = 1 for my test data.

Mol2Vec is used instead of other two approaches as the MSE and MAE obtained by Mol2Vec is much better and Mol2Vec uses the word embeddings which makes it more optimized

Hence Mol2Vec and SVR with hyperparameters set to CV = 100 and epsilon =1 produced the most optimised output.

How to run the program:

The paths are specified in the program itself. Those are absolute paths. So if one needs to run the program, he/she needs to change the paths in the program manually

And then run the program as a normal python file.