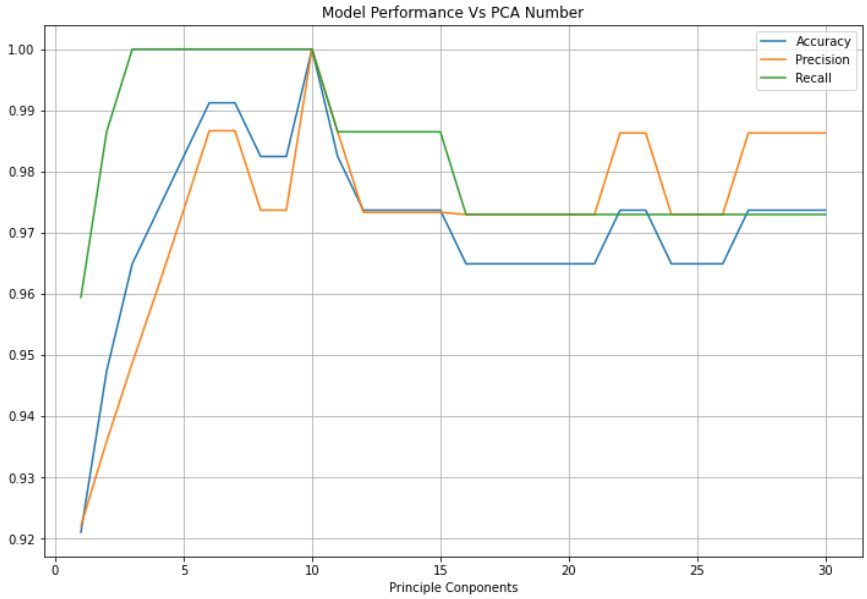
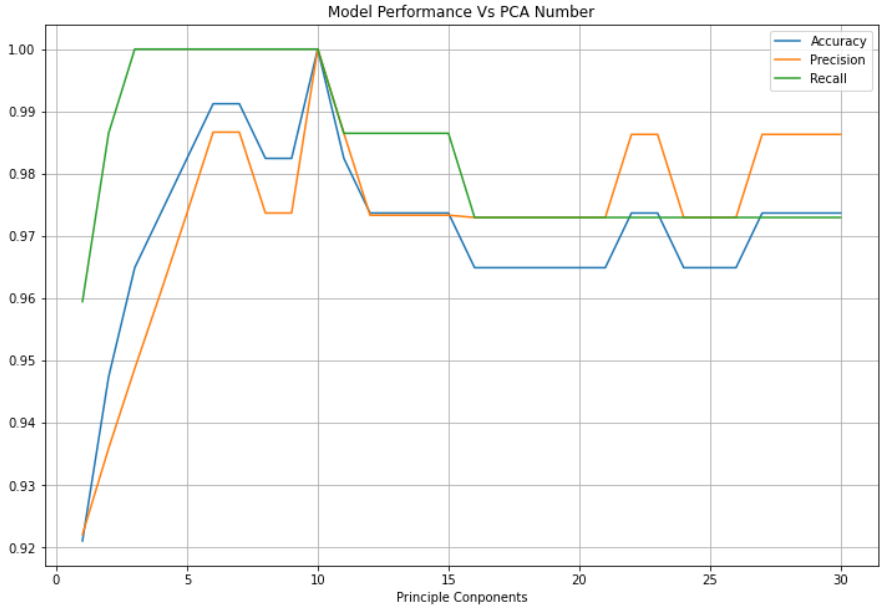
# Homework 4

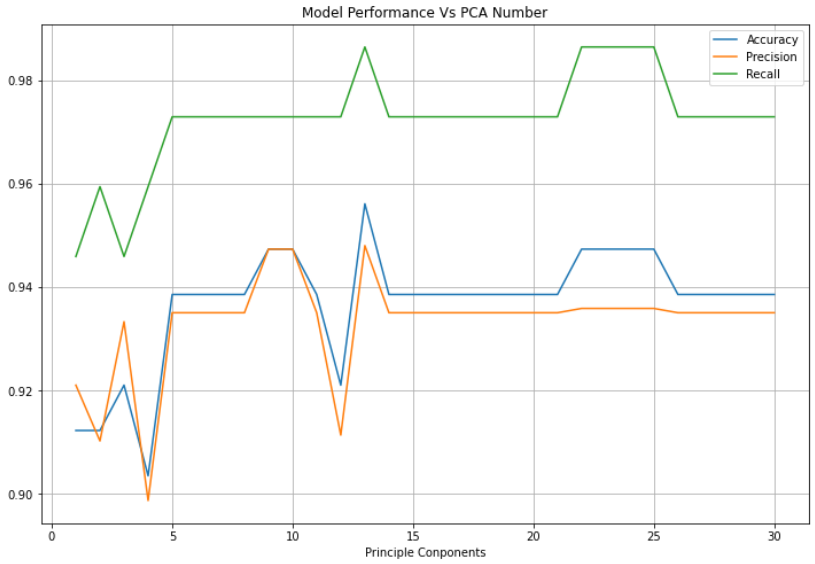
GitHub: https://github.com/pballou/ECGR_4105/tree/master/Homework/homework_4

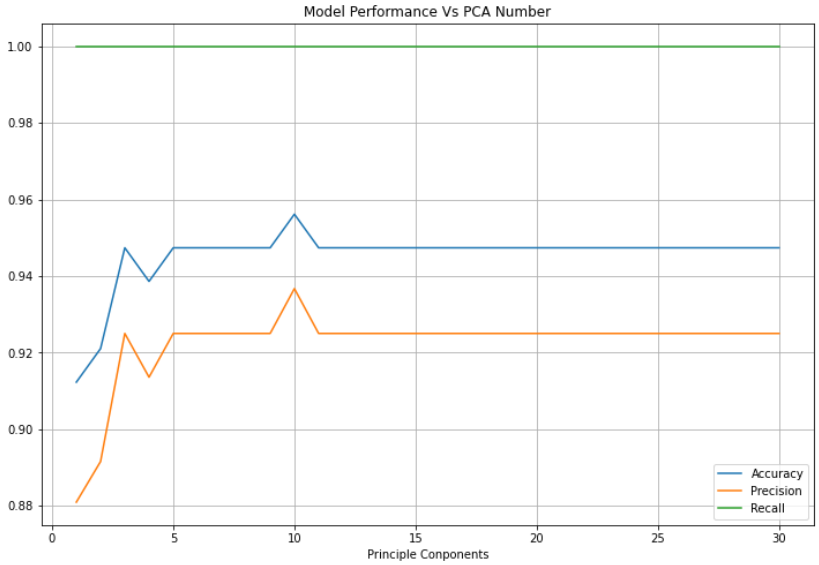
1. Use the cancer dataset to build an SVM classifier to classify the type of cancer.
    a. From my testing, the number of components that gives the best results is (K = 10). This is using min-max scaler, C (regularization parameter) of 10, and rbf as the kernel.
    b. Accuracy, precision, and recall over 30 K's.

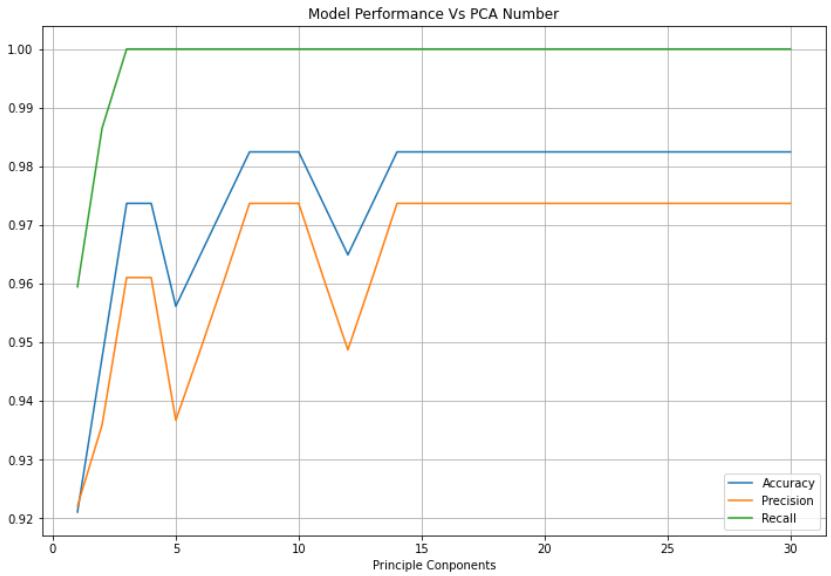



    c. Plot and compare the accuracies for different kernels. In order: rbf, sigmoid, poly, linear.



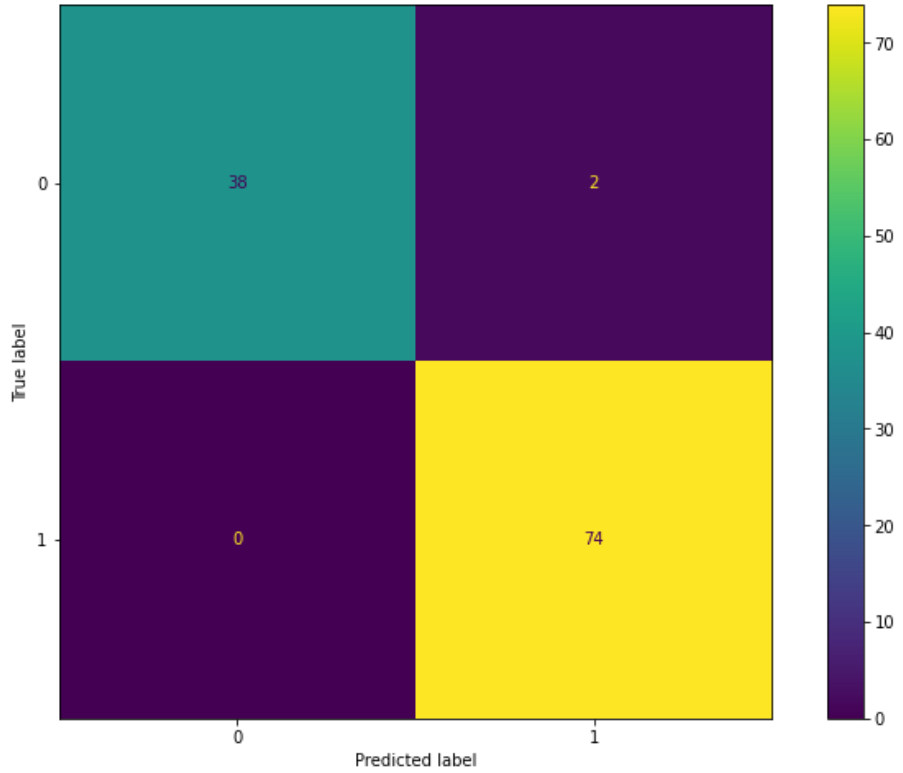

Model Performance Vs PCA Number
Accuracy
Precision
Recall
0.98
0.96
0.94
0.92
0.90
0
5
10
15
20
25
30
Principle Conponents


Model Performance Vs PCA Number
1.00
0.98
0.96
0.94
0.92
0.90
0.88
Accuracy
Precision
Recall
0
5
10
15
20
25
30
Principle Conponents


Model Performance Vs PCA Number
1.00
0.99
0.98
0.97
0.96
0.95
0.94
0.93
0.92
Accuracy
Precision
Recall
0
5
10
15
20
25
30
Principle Conponents

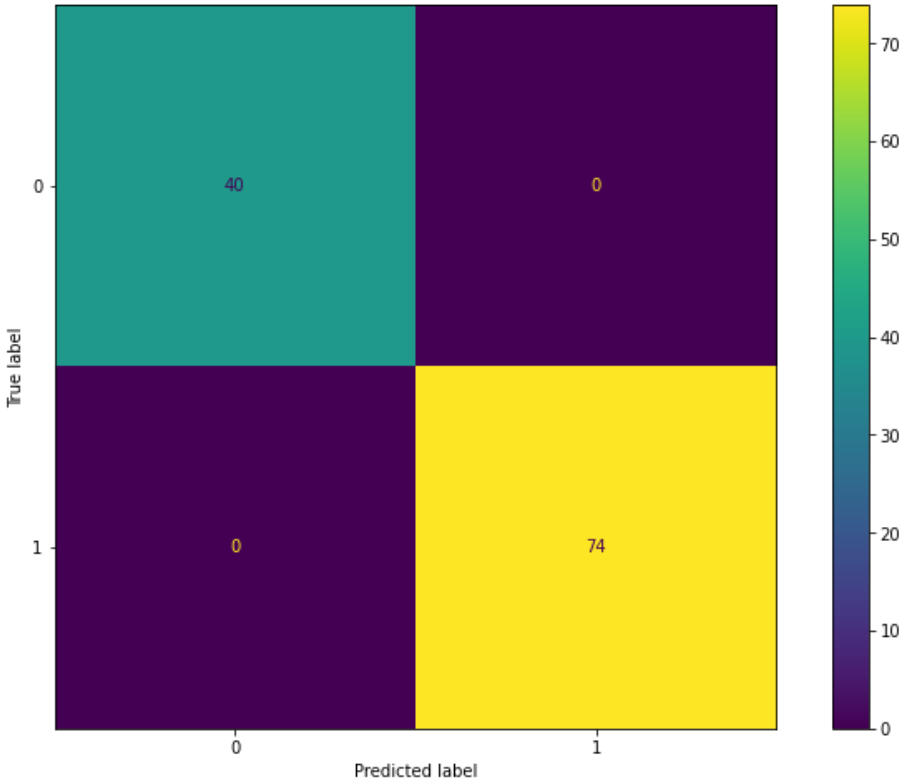
d. Compare your results against the logistic regression that you have done in homework 3.
   i. Homework 3 results:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 1.00 | 0.95 | 0.97 | 40 |
| 1 | 0.97 | 1.00 | 0.99 | 74 |
| accuracy | | | 0.98 | 114 |
| macro avg | 0.99 | 0.97 | 0.98 | 114 |
| weighted avg | 0.98 | 0.98 | 0.98 | 114 |




   ii. Homework 4 results:

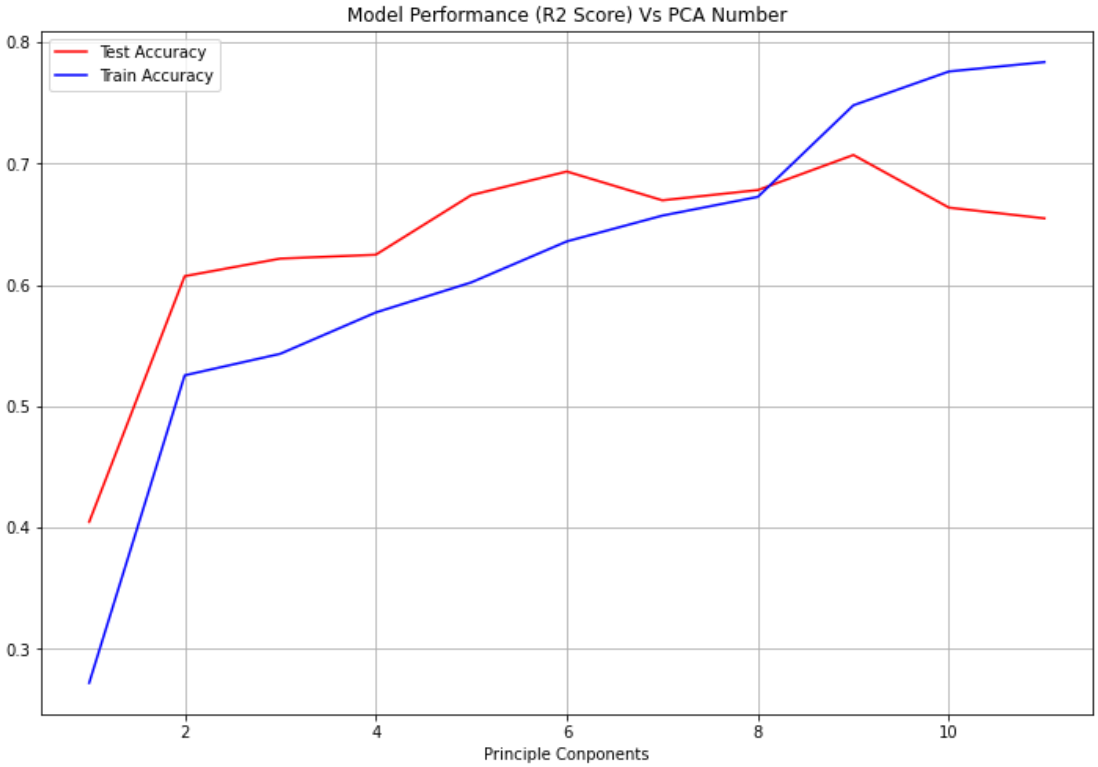
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0.0 | 1.00 | 1.00 | 1.00 | 40 |
| 1.0 | 1.00 | 1.00 | 1.00 | 74 |
| accuracy | | | 1.00 | 114 |
| macro avg | 1.00 | 1.00 | 1.00 | 114 |
| weighted avg | 1.00 | 1.00 | 1.00 | 114 |

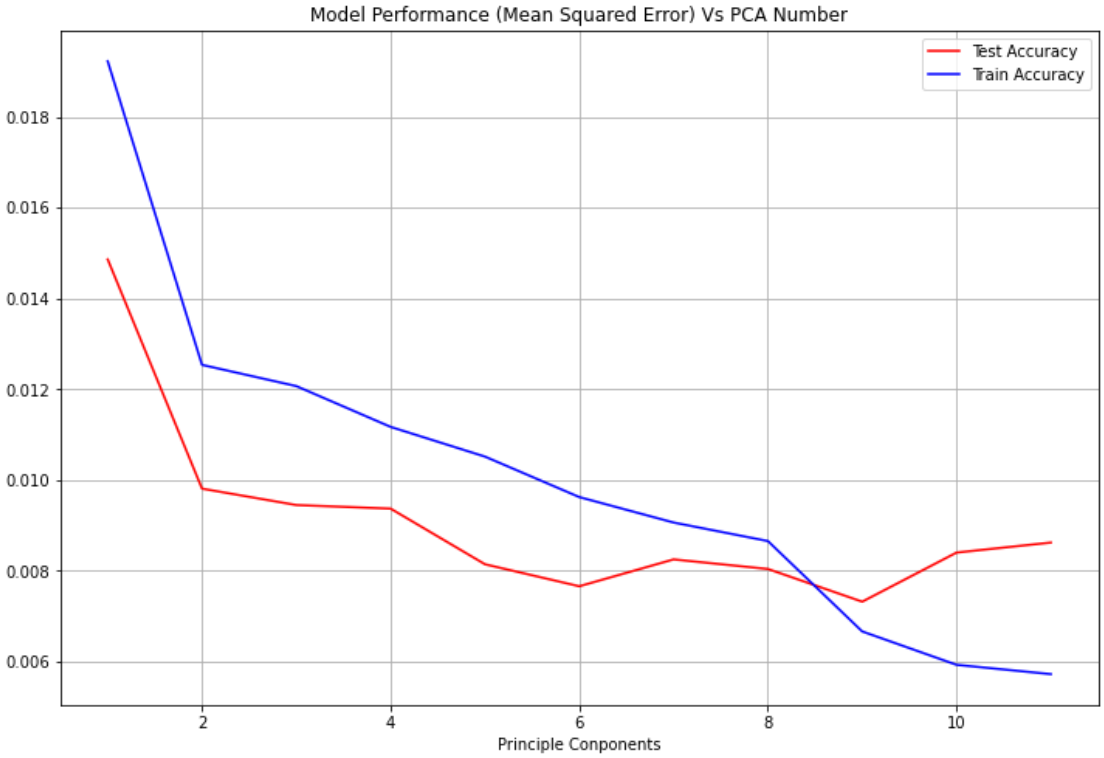



- While homework 3 was able to predict very well, the SVC that we used in this homework was much more accurate than the logistic regression from homework 3.

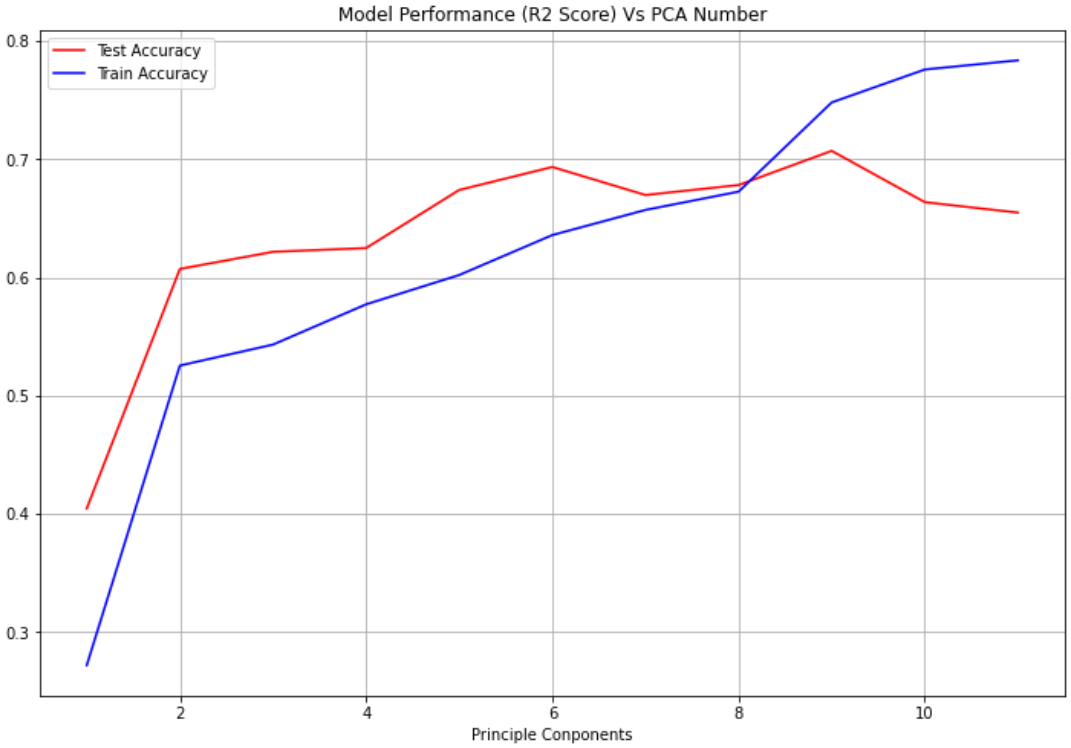
2. Use SVR model to predict housing price.
    a. Compare your results against linear regression with regularization loss that you already did in homework1.
        i. Homework 1 results:
        Min max training cost part b: 0.001434780826566402
        Min max testing cost part b: 0.0015886939687538653
        ii. Homework 4 results:
        Train set MSE: 0.006657826026871647
        Test set MSE: 0.0073137120497441905
        Train set R2 score: 0.7479439736466951
        Test set R2 score: 0.7069932866645069
        - Comparing the results, the linear regression from homework seems to have performed better than the SVR we used in this homework. From what I understand, the SVR should be a better model than linear regression, so I think my results in homework 1 were incorrect for some reason, but I am not sure why.

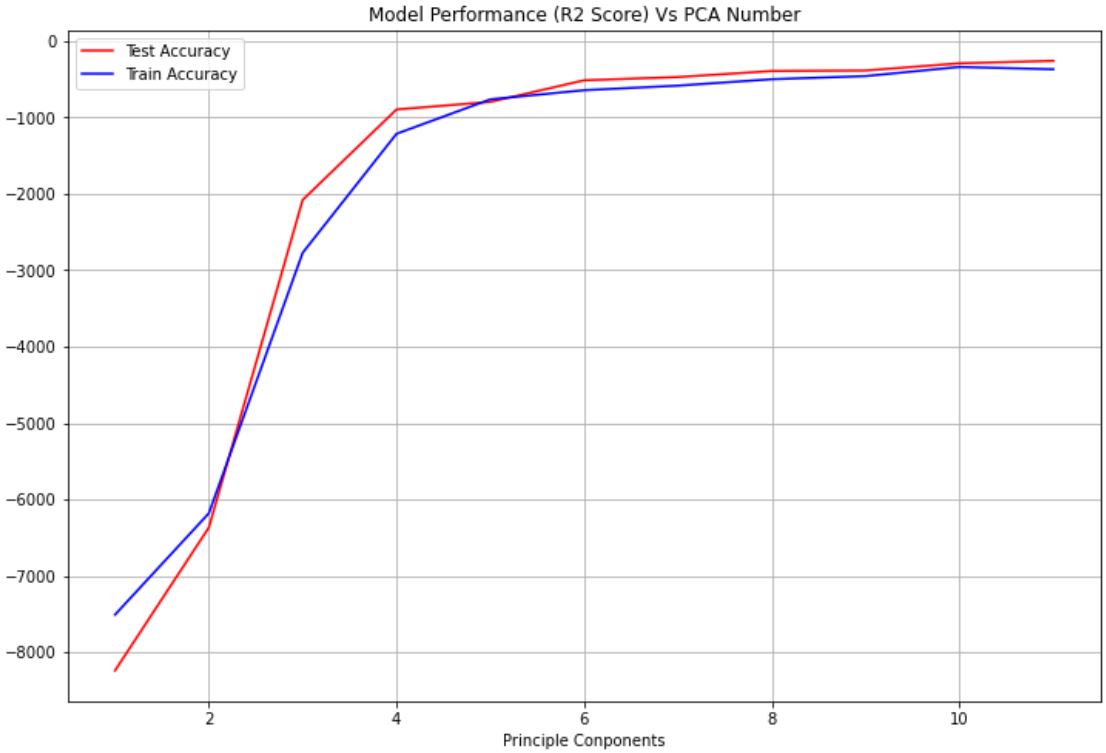
b. From my testing, the number of components that gives the best results is (K = 9). This is using min-max scaler, C (regularization parameter) of 1, and rbf as the kernel.

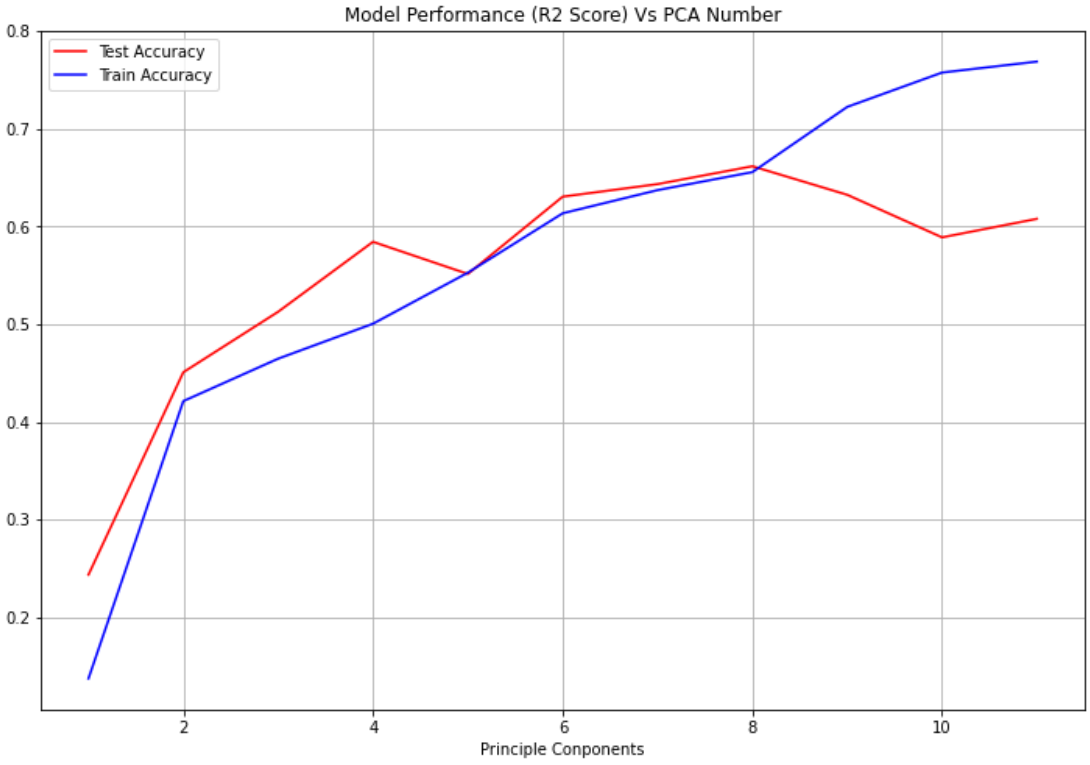


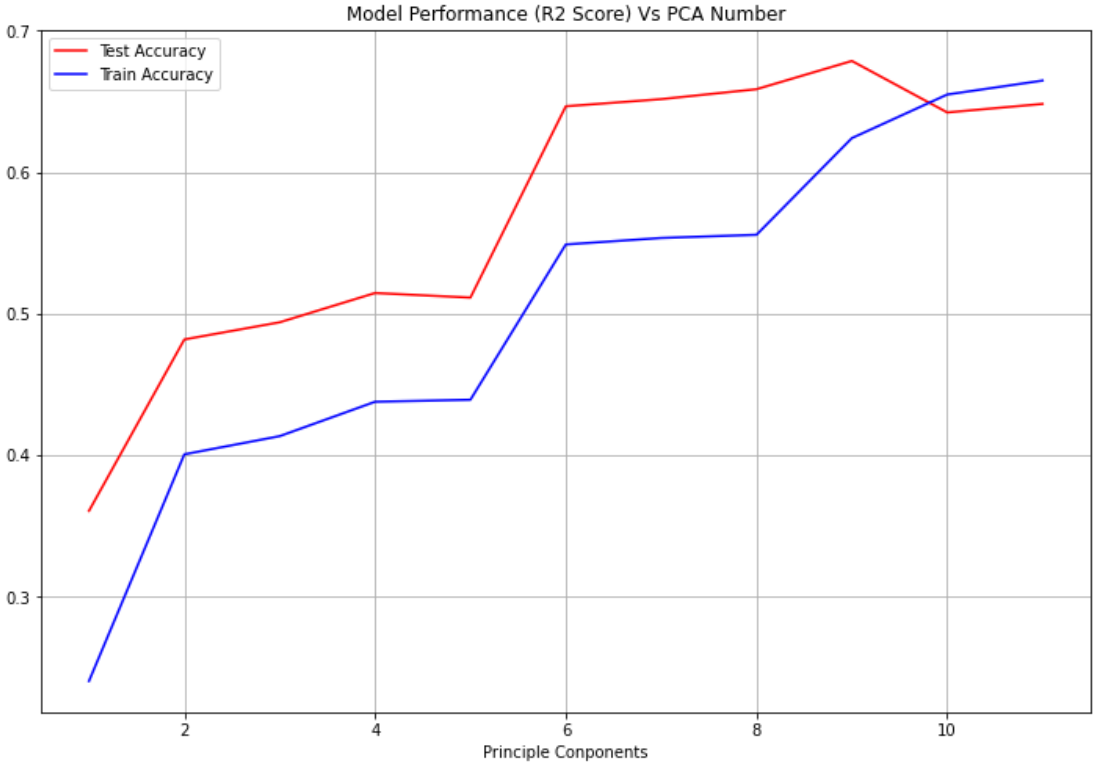



c. Plot and compare the accuracies for different kernels. In order: rbf, sigmoid, poly, linear.

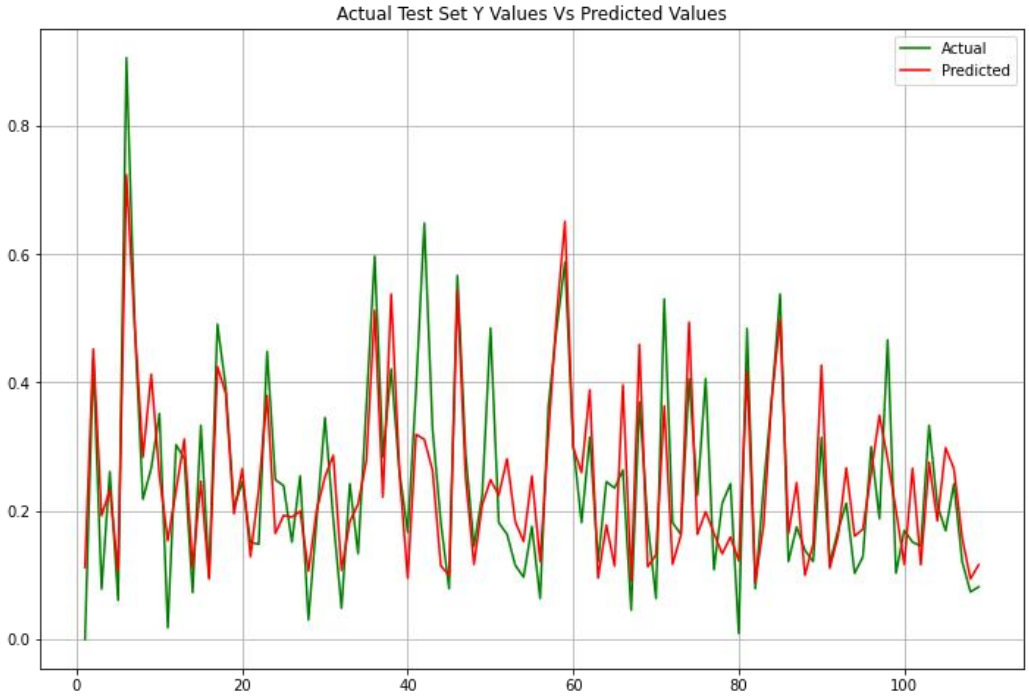





- As we can see, the rbf kernel performs best. This graph plots the actual Y values from the test set vs the predicted values from the final model.


Actual Test Set Y Values Vs Predicted Values
Actual
Predicted
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100