Homework 4
ECGR-4105-001
Joshua Foster, 801268118
Repo Link: https://github.com/jfoste81/ECGR4105_Homework/tree/main

Problem 1:

I experimented with the linear, poly, RBF, and sigmoid kernels (results charted below in confusion matrices and compared using a bar graph) and found RBF to have the best performance (98% accuracy), with linear being a close second (97%). I would say, overall, the performance between the two top performers is comparable as they made the same number of false negatives, which is the worst-case scenario in the context of cancer data. After the top two kernels, I would say that there is a fairly steep drop off until the third best kernel, the sigmoid kernel, and then an even further drop off to the worst performing kernel, the poly kernel. The sigmoid kernel achieved 93% accuracy, while the poly kernel achieved 91% accuracy. We see slightly more false negatives and significantly more false positives with the sigmoid kernel. Surprisingly, the poly kernel had zero false positives but had a dreadful 10 false negatives. This insane number of false negatives (10/42 possible positive cases) is a terrible performance given the overall goal of accurately detecting cancer. While false positives are not great, it is infinitely better than missing the cancer entirely and giving it a chance to progress undetected.

Below the bar graph, we have the results of my logistic regression from homework 3. The logistic regression from homework 3 outperforms 3 of the 4 kernels I experimented with, even before Regularization (which notably did not increase performance in the model), but shares equivalent performance with the rbf kernel. They have the exact same predictive accuracy and, on this dataset, have the exact same performance.
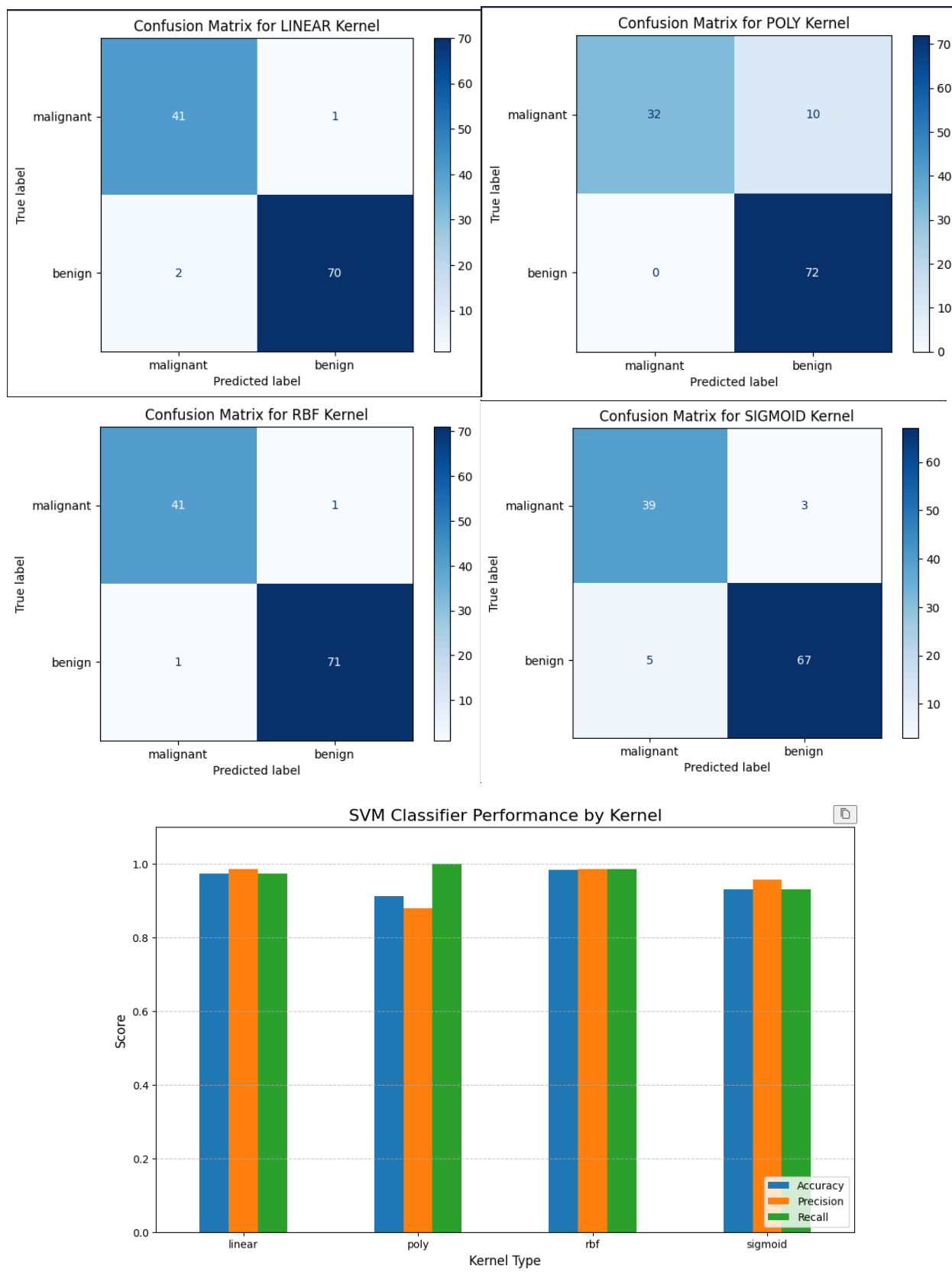
Figure 1: Overview of all Kernel performance

```
Accuracy: 0.98
Precision: 0.99
Recall: 0.99
F1 Score: 0.99
```


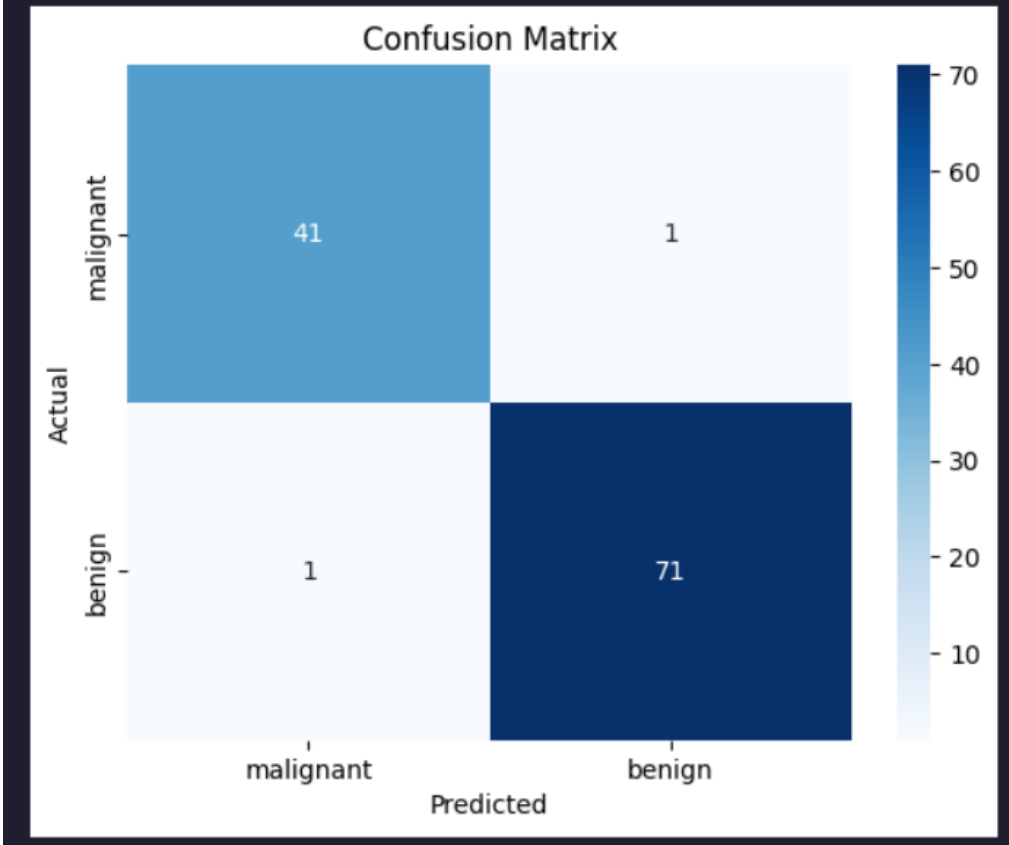*Figure 2: Logistic Regression from HW3*

Problem 2:

For this problem, I used the linear, RBF, and poly kernels and implemented a simple Ridge model (which is a linear regression model with L2 regularization, ie homework1). Below are the charted R^2 (accuracy) scores and the plotted predictions of the linear and RBF kernels and Ridge model. We can see that the Ridge model outperformed the SVR across the board on the housing dataset. The best performing SVR, by far, was the linear kernel with a .6180 R^2, followed by the RBF kernel with a .5970 R^2, and ending with the poly kernel that scored a dismal .5415 R^2. All of these were lower than the Ridge model's .6436 R^2, but interestingly the linear kernel was the closest in performance. I find it interesting that both linear-related models performed the best while the poly kernel performed the worst.
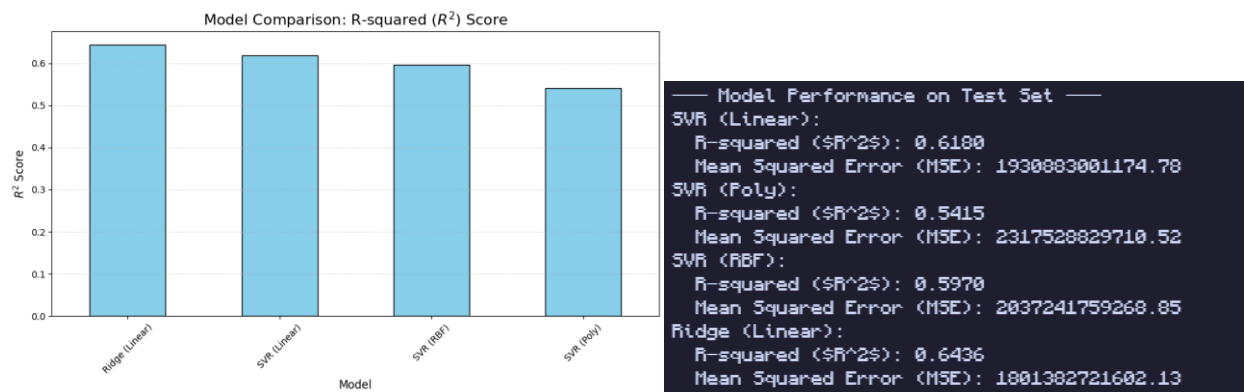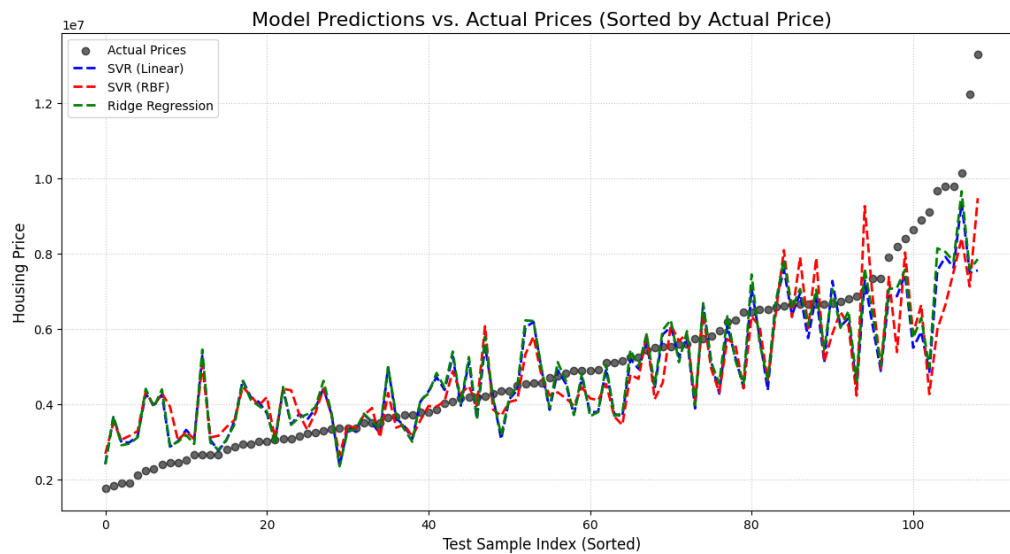

Figure 3: Accuracies for Models

Figure 4: Graph of Homework 4 Models