Sean Timkovich-Camp
Garrett Holtz

Assignment 3 Lab Report

## Regression:

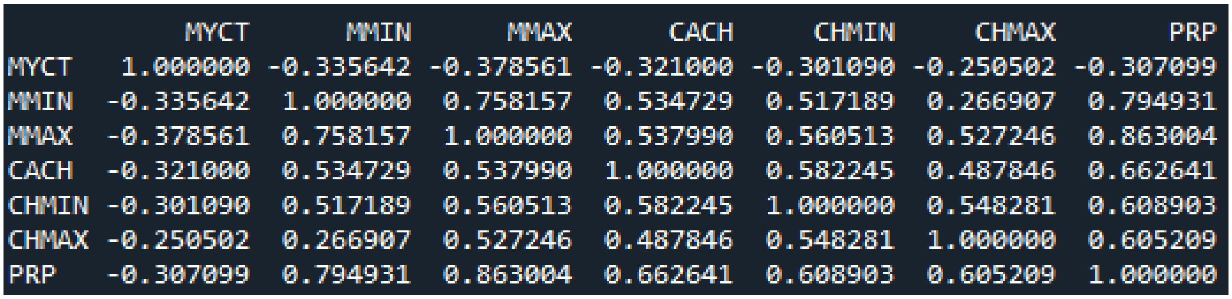
For our regression testing, we used the Computer Hardware dataset from https://archive.ics.uci.edu/ml/datasets/Computer+Hardware.

| | |
|---|---|
| Number of Features | 7 |
| Names of the features | MYCT, MMIN, MMAX, CACH, CHMIN, CHMAX, PRP |
| Name of target | ERP |
| Number of samples | 209 |
| Description of data | (209, 8) |

First five rows of data:

| Row # | MYCT | MMIN | MMAX | CACH | CHMIN | CHMAX | PRP |
|---|---|---|---|---|---|---|---|
| 1 | 125 | 256 | 6000 | 256 | 16 | 128 | 198 |
| 2 | 29 | 8000 | 32000 | 32 | 8 | 32 | 269 |
| 3 | 29 | 8000 | 32000 | 32 | 8 | 32 | 220 |
| 4 | 29 | 8000 | 32000 | 32 | 8 | 32 | 172 |
| 5 | 29 | 8000 | 32000 | 32 | 8 | 16 | 132 |

Correlation between the features:



```
            MYCT      MMIN      MMAX      CACH     CHMIN     CHMAX       PRP
MYCT   1.000000 -0.335642 -0.378561 -0.321000 -0.301090 -0.250502 -0.307099
MMIN  -0.335642  1.000000  0.758157  0.534729  0.517189  0.266907  0.794931
MMAX  -0.378561  0.758157  1.000000  0.537990  0.560513  0.527246  0.863004
CACH  -0.321000  0.534729  0.537990  1.000000  0.582245  0.487846  0.662641
CHMIN -0.301090  0.517189  0.560513  0.582245  1.000000  0.548281  0.608903
CHMAX -0.250502  0.266907  0.527246  0.487846  0.548281  1.000000  0.605209
PRP   -0.307099  0.794931  0.863004  0.662641  0.608903  0.605209  1.000000
```

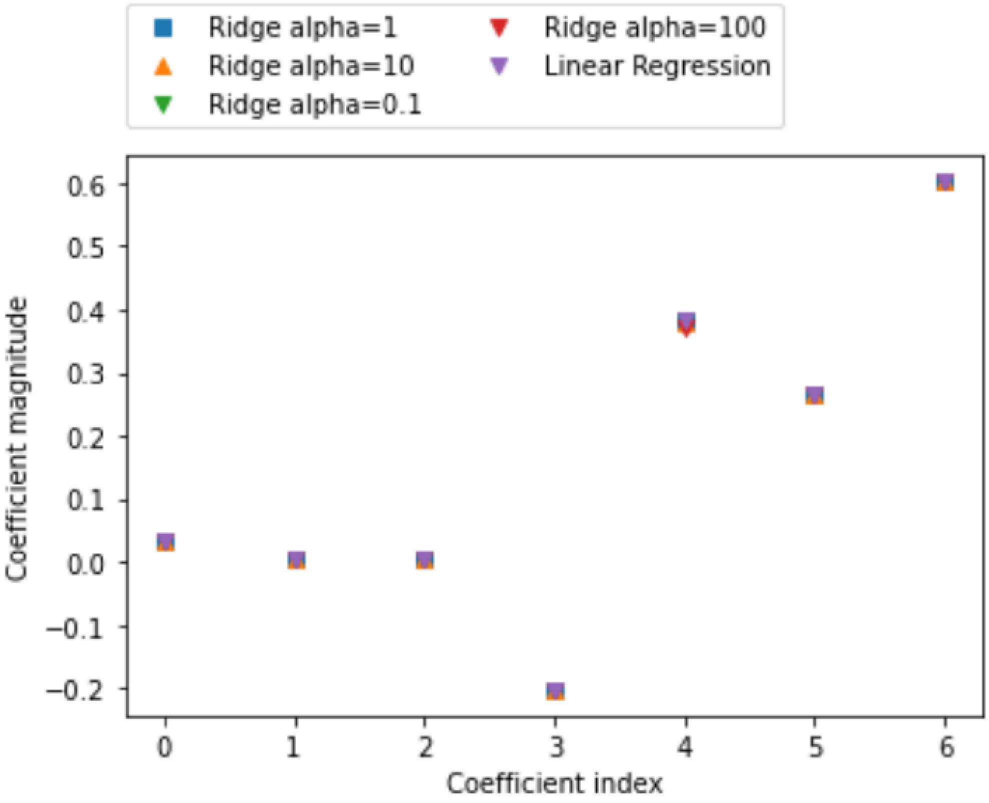
Best Parameters:
Using the GridSearchCV function of sklearn, we able to automatically try multiple parameters of alpha. We were able to deduce the optimal alpha for both Lasso and Ridge Regression.
Random_state = 43

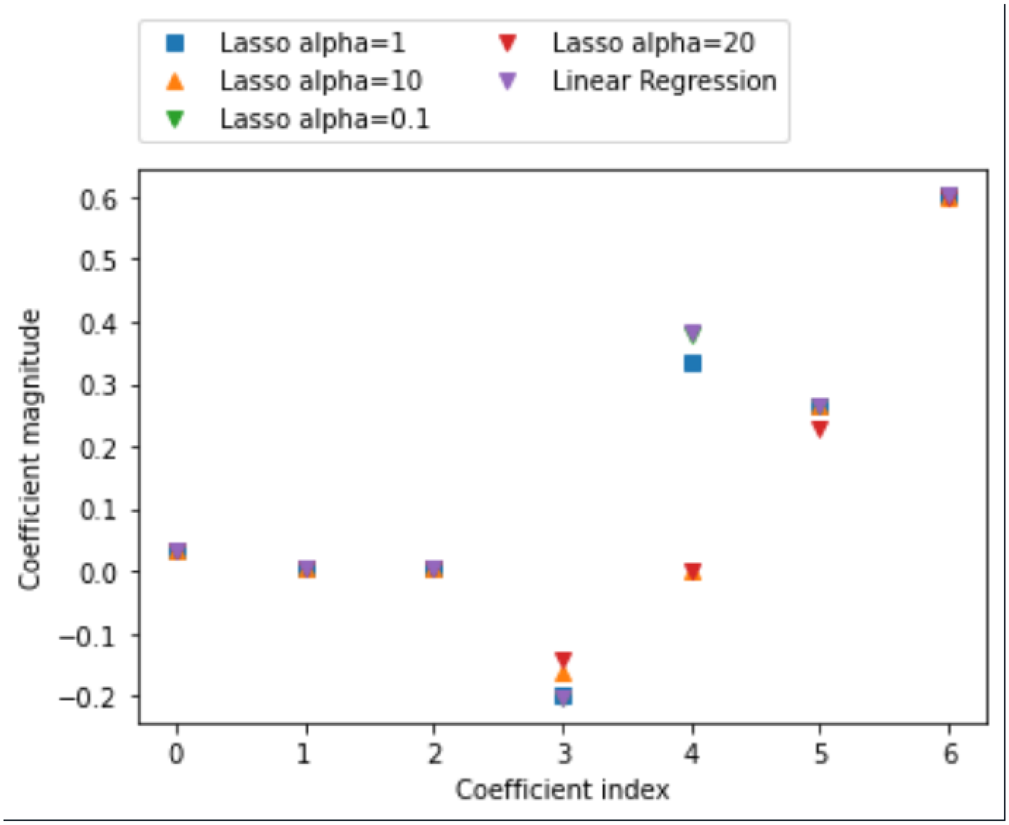
| | |
|---|---|
| Ridge | 100 |
| Lasso | 20 |

After finding the optimal parameter alpha, we used that parameter to predict the target data. We used $R^2$ and RMSE as our scoring methods to measure the accuracy of our methods.

| Method | $R^2$ | RMSE |
|---|---|---|
| Lasso | 0.8821 | 5.2389 |
| Ridge | 0.8821 | 5.2389 |
| Linear | 0.8820 | 5.2399 |

Comparison of Ridge, Lasso, and Linear Regression Models:
Using the optimal parameter that we found earlier, we made a linear model with calculated coefficients. The following figures show those coefficients and how they change with the parameter alpha.


■ Ridge alpha=1
▲ Ridge alpha=10
▼ Ridge alpha=0.1
▼ Ridge alpha=100
▼ Linear Regression
Coefficient magnitude
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
0
1
2
3
4
5
6
Coefficient index

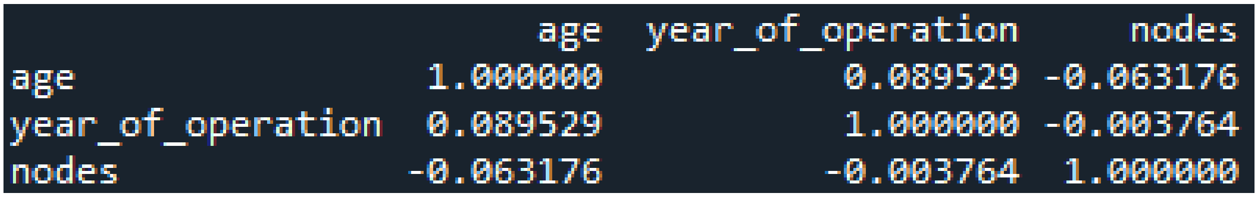

■ Lasso alpha=1
▲ Lasso alpha=10
▼ Lasso alpha=0.1
▼ Lasso alpha=20
▼ Linear Regression
Coefficient magnitude
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
0
1
2
3
4
5
6
Coefficient index

## Classification

For our classification testing, we used Haberman’s Survival dataset from https://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival.

| | |
|---|---|
| Number of Features | 3 |
| Names of the features | Age. Year of Operation, Nodes |
| Name of target | Survival Status |
| Number of samples | 306 |
| Description of data | (306, 4) |

Correlation between features:



```
                         age  year_of_operation     nodes
age                 1.000000           0.089529 -0.063176
year_of_operation   0.089529           1.000000 -0.003764
nodes              -0.063176          -0.003764  1.000000
```

After training our Logistic Regression model, we calculated the accuracy of our model to be 75.8%.

Accuracy of the Logistic Regression: 0.75806

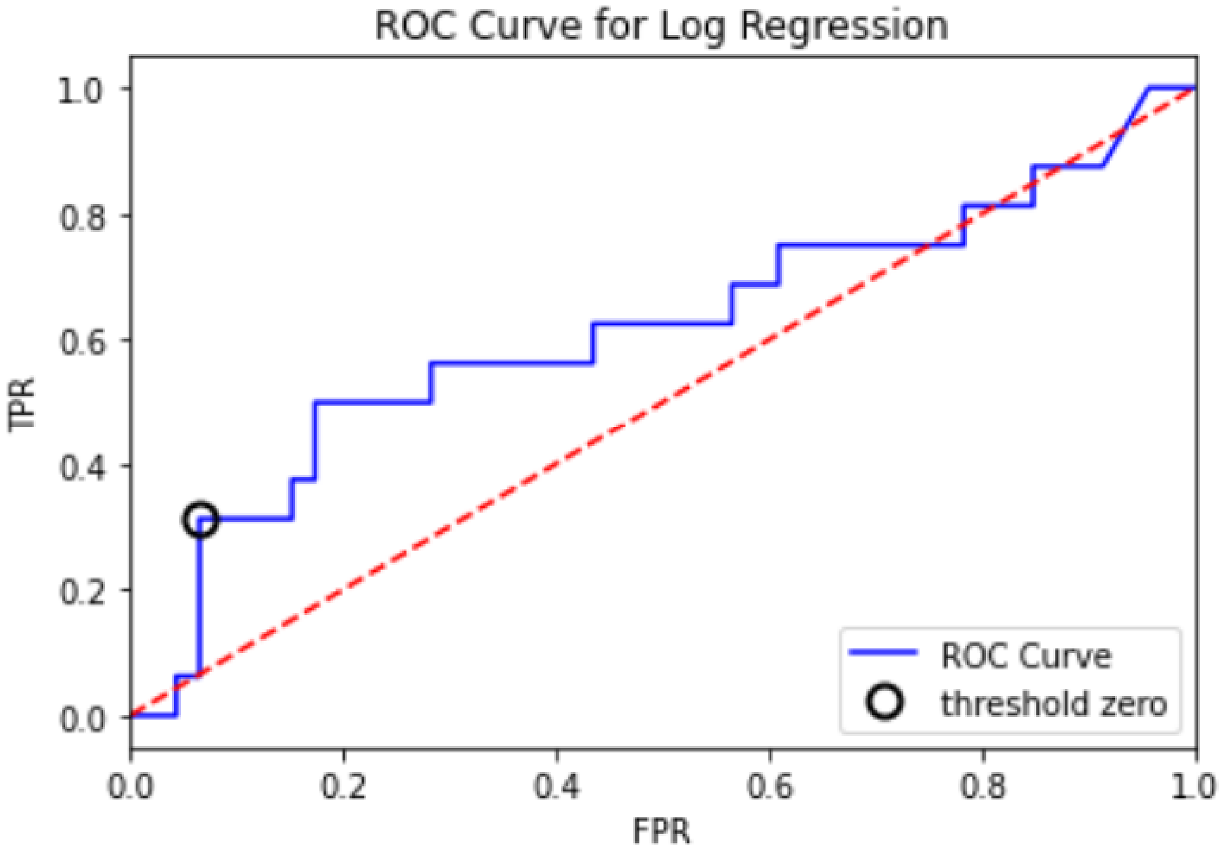
With our model, we ran a prediction of the test target data using the test data. With the predictions, we were able to create a confusion matrix and solve for precision, recall, sensitivity, and accuracy, as seen below.

Model Performance:

| | |
|---|---|
| Precision | 0.9347 |
| Recall | 0.7818 |
| Sensitivity | 0.7818 |
| Accuracy | 0.7580 |

ROC Curve:

To determine the effectiveness of the model, we plotted a ROC Curve. From our data, we can see our model is not performing the best it could. There could be a model better suited to this data.


ROC Curve for Log Regression
TPR
FPR
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
ROC Curve
threshold zero