ISYE 8803 HDDA

Homework 4

July 13th, 2021

2.

Provided are the features used for the question:

Number	Description
1	log_gdppc
2	log_pop
3	age_1
4	age_2
5	age_3
6	age_4
7	age_5
8	educ
9	age_median

a)

$\lambda_{ridge} = 0.6018896$

```
> coef.ridge
[1]  0.053461206 -0.007220268 -0.278585915 -0.093601429
0.279340848  0.708965801  0.613538358  0.018237348  0.004080870
```

From the results of our coefficient output, we can see that at the optimal lambda, the features weighted with importance are the 3, 5, 6 and 7th features corresponding to, all others have been pushed to 0.

Which in our case means that age_1, age_3, age_4 and age_5 were deemed variables of higher importance / correlation in comparison to , and all others were driven closer to 0. I would also assume that the age_median variable may cause some multicollinearity issues which ridge may be able to address.

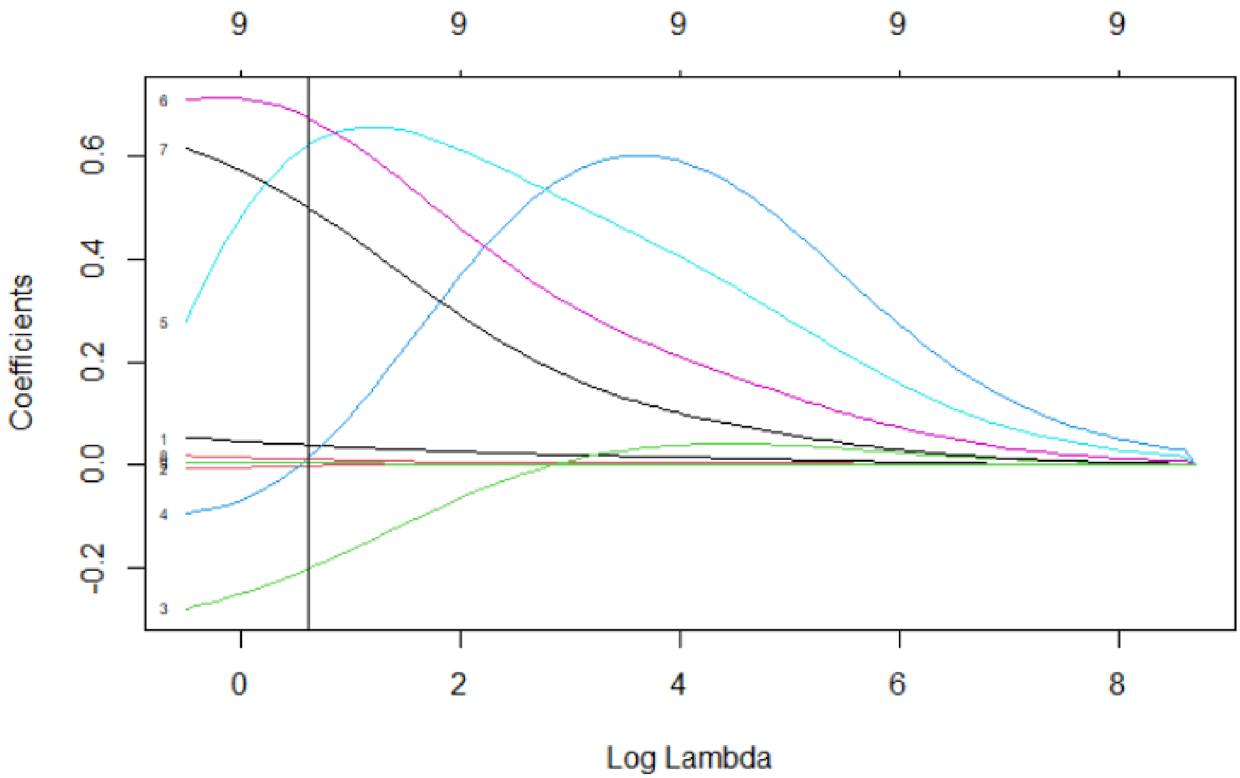

Figure 1: Ridge Regression Coefficients Plot

$MSE_{ridge} = 0.1608326$

b)

Lasso Regression

$\lambda_{lasso} = 0.00115437$
$\log(\lambda_{lasso}) = -6.7642$

```
> coef.lasso
[1]  0.100045255 -0.008731155 -0.448847903 -0.100778816
-1.308575752  0.615425560  0.539299641  0.034986872  0.000000000
```

From the results of our coefficient plot, we can see that Lasso seems to set the 9th coefficient to 0, and the remaining coefficients are still being used by the model. In our case the log lambda is being used. We can see that the results are fairly consistent with Ridge, in that coefficients 3-7 are deemed useful and the remaining coefficients are very close to 0.

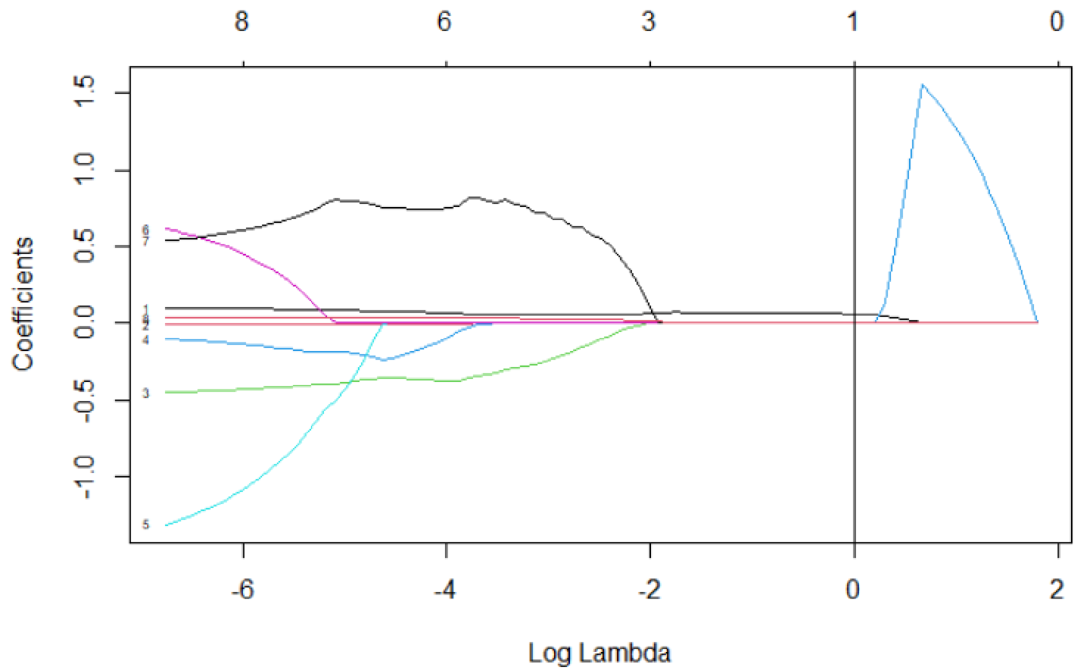

Figure 2: Lasso Regression Coefficients Plot

$MSE_{lasso} = 0.1813994$

c)

$\lambda_{OLS\ Ridge} = 1.00044$

```
> coef.alasso2
[1]  0.044634499  0.000000000 -0.524147480  0.625942583
-0.002441487  0.273682950  2.713964381  0.000000000  0.000000000
```

We can see from the OLS estimator used for ridge that the 2nd, 8 and 9th coefficients are deselected from the model, and the 3rd to 7th variables seem to be performing well in comparison to the previous models used. We can also see that certain features of higher magnitude are not being penalized with a constant compared to lasso, as we seem much larger spikes for the 6th and 7th variables.

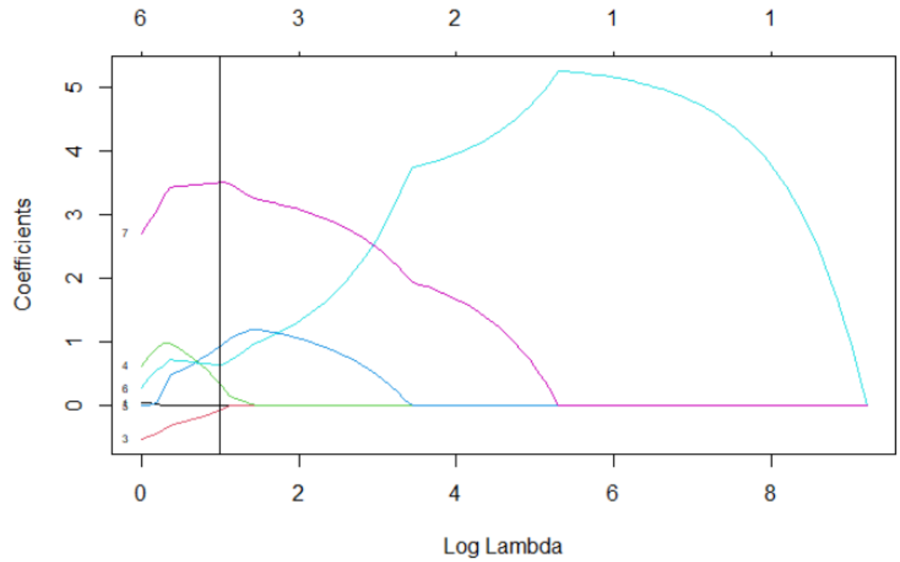

Figure 3: Adaptive Lasso Ridge

$MSE_{alasso} = 0.170724$

d)
Elastic Net

We can see from the plot below that the lambda.min values over each iteration are slowly decreasing and resulting in an optimal value of 0.99 being selected. This means that the model is performing better with L1 regularlization with lasso compared to ridge regression.

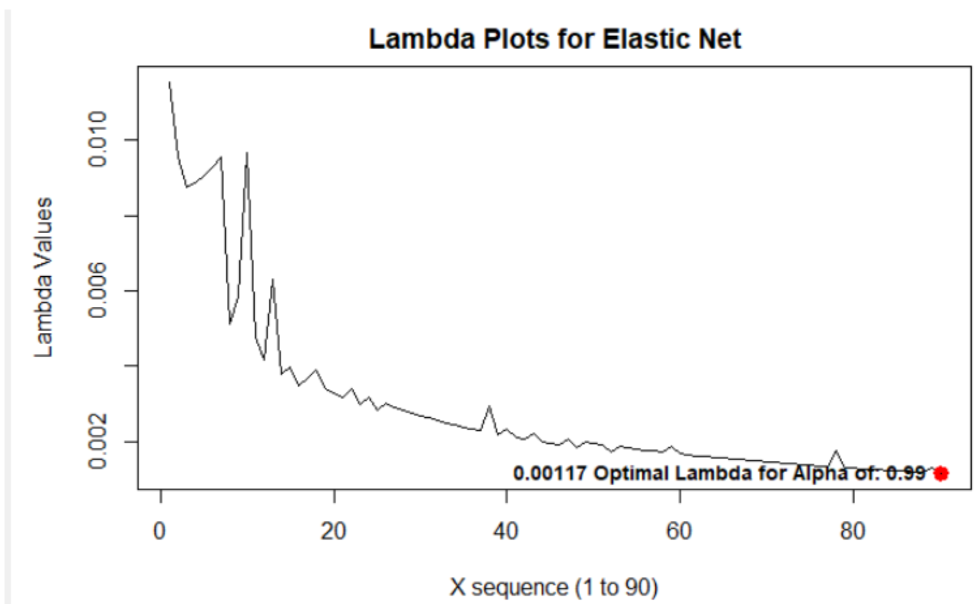

Figure 4: Alpha Calculation Elastic Net

$alpha = 0.99$

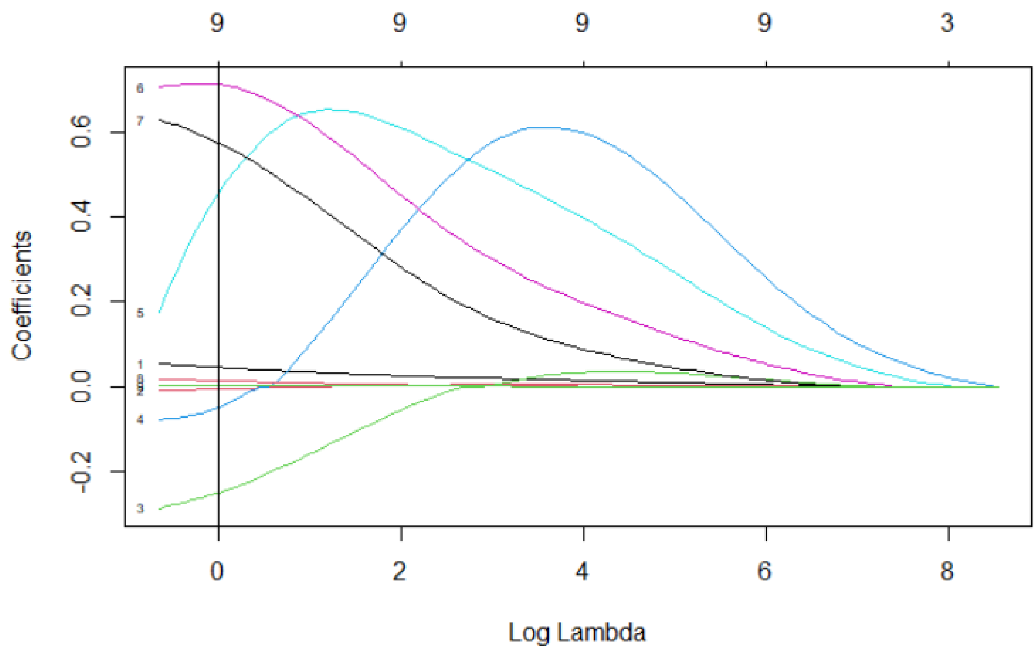

Figure 5: Elastic Net Regression Coefficients Plots

$MSE_{elastic\ net} = 0.1622157$

2d)

We can see from the outputs that the MSE's are the following:

$MSE_{ridge} = 0.1608326$
$MSE_{lasso} = 0.1813994$
$MSE_{alasso} = 0.170724$
$MSE_{elastic\ net} = 0.1622157$

From these results we can conclude that the ridge regression is likely the better candidate due to the lowest MSE compares to the others. Choosing ridge also lines up well with our data since we're likely experiencing multicollinearity with the age variables and the age_median variable, which ridge handles better compared to lasso.

We can support this statement since the lasso regression deselects the age_median variable from the model, as it yields no predictive value.