

Question – 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer:

Optimal value of alpha for Ridge regression – 0.11

Optimal value of alpha for Lasso regression – 0.3

The model performance for this values

	Metric	Linear Regression	Lasso Regression	Ridge Regression
0	R2 Score (Train)	0.896954	0.780393	0.896919
1	R2 Score (Test)	0.789073	0.792312	0.795031

When you double the alpha value, model regularization is increased. The model becomes underfit.

The important predictor variables

```
array([-4.46282365e-04, -0.00000000e+00, -0.00000000e+00,  1.57551445e-06,
       -0.00000000e+00,  0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
       -0.00000000e+00, -0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
        0.00000000e+00,  0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  3.07691937e-03,  2.92578836e-03,
       -0.00000000e+00,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
       -0.00000000e+00,  6.17757328e-05, -0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  0.00000000e+00, -0.00000000e+00,
        0.00000000e+00,  3.96546874e-05, -0.00000000e+00,  2.95294606e-05,
       -4.54217580e-08,  1.11107563e-04, -0.00000000e+00, -0.00000000e+00,
       -0.00000000e+00,  0.00000000e+00,  1.17391606e-04,  1.34134642e-04,
       -0.00000000e+00,  1.76514123e-04,  0.00000000e+00,  0.00000000e+00,
        0.00000000e+00, -0.00000000e+00, -0.00000000e+00,  0.00000000e+00,
        0.00000000e+00,  0.00000000e+00,  0.00000000e+00, -0.00000000e+00,
        0.00000000e+00, -0.00000000e+00,  0.00000000e+00,  3.29109464e-04,
        0.00000000e+00,  0.00000000e+00, -0.00000000e+00,  2.08150338e-04,
        0.00000000e+00,  0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
        0.00000000e+00, -0.00000000e+00,  0.00000000e+00,  0.00000000e+00])
```

The ones that are higher in the values like greater than zero are the most important predictor variables

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Answer :

Lasso regression performed better than Ridge regression in terms of reducing the difference between train and testing model performance

It also helps in feature elimination by reducing the number of predictor variables

	Metric	Linear Regression	Lasso Regression	Ridge Regression
0	R2 Score (Train)	0.896954	0.780393	0.896919
1	R2 Score (Test)	0.789073	0.792312	0.795031

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Answer

We can remove -4.4, -4.5, any any variable with coefficient as zero

```
array([ -4.46282365e-04, -0.00000000e+00, -0.00000000e+00,  1.57551445e-06,
        -0.00000000e+00,  0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
        -0.00000000e+00, -0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
         0.00000000e+00,  0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
         0.00000000e+00,  0.00000000e+00,  3.07691937e-03,  2.92578836e-03,
        -0.00000000e+00,  0.00000000e+00,  0.00000000e+00,  0.00000000e+00,
        -0.00000000e+00,  6.17757328e-05, -0.00000000e+00,  0.00000000e+00,
         0.00000000e+00,  0.00000000e+00,  0.00000000e+00, -0.00000000e+00,
         0.00000000e+00,  3.96546874e-05, -0.00000000e+00,  2.95294606e-05,
        -4.54217580e-08,  1.11107563e-04, -0.00000000e+00, -0.00000000e+00,
        -0.00000000e+00,  0.00000000e+00,  1.17391606e-04,  1.34134642e-04,
        -0.00000000e+00,  1.76514123e-04,  0.00000000e+00,  0.00000000e+00,
         0.00000000e+00, -0.00000000e+00, -0.00000000e+00,  0.00000000e+00,
         0.00000000e+00,  0.00000000e+00,  0.00000000e+00, -0.00000000e+00,
         0.00000000e+00, -0.00000000e+00,  0.00000000e+00,  3.29109464e-04,
         0.00000000e+00,  0.00000000e+00, -0.00000000e+00,  2.08150338e-04,
         0.00000000e+00,  0.00000000e+00, -0.00000000e+00, -0.00000000e+00,
         0.00000000e+00, -0.00000000e+00,  0.00000000e+00,  0.00000000e+00])
```

Question - 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answer

To make the model more robust and generalizable, the model shouldn't be overfit or underfit.

The model should work perfectly for unseen similar data. For this the model complexity and variance needs to be balanced. The total error needs to be minimum

Regularization helps in managing model complexity thus avoids overfitting