

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 is – 20

Lasso is – 100

On doubling the alpha values –

1. Ridge

Alpha=20

Alpha=40

```
#Fitting Ridge model for alpha = 20 and printing coefficients whi
alpha = 20
ridge = Ridge(alpha=alpha)

ridge.fit(X_train, y_train)
print(ridge.coef_)
```

```
[-1.32001711e+02  7.15361470e+01  4.35422786e-01  2.71525886e+03
-2.16454840e+03 -1.69633907e+03  0.00000000e+00  3.03737965e+03
 9.86188202e+03  5.74147469e+03  2.43402559e+02  1.08496617e+01
 2.46989704e+01  6.39507834e+03 -5.09617532e+03  2.09726614e+01
 5.27558199e+00 -2.00529166e+00  2.42429518e+01 -4.83348159e+01
-1.95915570e+03  1.90094025e+01  2.97784874e+01 -1.40595982e+01
 3.47282919e+01  1.07542105e+03 -3.39968334e+03  5.43032027e+02
 1.57998268e+03 -4.73417044e+03 -7.67478954e+03  5.22449162e+03
 9.11932346e+02  4.77803124e+03  4.55839621e+03  6.97995029e+01
 5.64359288e+03  7.80939445e+00  1.55279430e+01  2.86985397e+01
-4.76421280e+00  2.59513470e+01  3.08079544e+01  1.02695884e+02
-1.48125770e+00  2.09033666e+02  3.36434473e+02  3.32007216e+03
 3.19581016e+03 -6.13738590e+02  3.18841790e+03 -1.30938503e+03
 1.61350529e+03 -9.71836526e+01  5.59805252e+03 -1.23908892e+03
-1.74637500e+03  7.47718226e+02  8.18271310e+02  1.22803665e+03
 6.14625231e+03 -4.69883080e+03 -4.56331970e+03  9.27731233e+03
-3.58395133e+03 -4.33627647e+03 -3.97708795e-02 -3.27612728e+02
-6.79607508e+03 -4.93362692e+03  3.19459804e+03 -8.01428078e+03
 8.27222333e+03  1.10245017e+04 -2.78668662e+03 -1.54773720e+03
 1.74789109e+02 -3.56396895e+03  2.02304152e+03  1.14301886e+04
-5.42045425e+03  1.86001778e+03  1.62388674e+01  4.54397643e+03
 6.95623057e+02  3.54604203e+02 -4.78077220e+03  2.77794975e+03
-8.83941159e+02 -8.75462684e+01 -1.26197963e+03 -1.22390427e+03
 1.15958193e+03  0.00000000e+00  0.00000000e+00 -1.22030952e+02
 9.45437088e+02  1.53315362e+03 -4.73181701e+03 -3.29042885e+03
 1.41747857e+03  2.39224439e+03  2.42606742e+03  1.80216427e+03
-2.85255830e+02 -8.26410155e+02  3.47907143e+03  5.11320953e+03
-3.37690089e+02  1.41174670e+03  1.33014420e+03  9.70765878e+02
 3.79429751e+02  1.05698614e+03  0.00000000e+00  1.40094843e+02
-4.58926127e+02 -5.27402240e+03 -1.16518526e+03  5.70105279e+03
-9.11879409e+02  9.22024275e+02  6.90546115e+03 -6.08285436e+01
 3.01918980e+03 -2.01077413e+03 -4.59699725e+02  3.96078351e+00
-3.71404809e+03 -8.11884922e+02  7.18428132e+02 -7.65040108e+02
-4.06425514e+03  6.86966500e+02 -1.19372677e+03  2.20598558e+03
 9.71786122e+02 -6.08285436e+01  1.21536396e+03 -3.36581023e+03
 6.03296845e+03  9.13065077e+01 -5.32959207e+02 -5.12309589e+03
 9.17634826e+02 -1.22867683e+03 -1.39478289e+02  6.47214854e+02
-2.08204333e+03 -1.34255585e+03  4.16524577e+03  2.68337334e+03
 7.76031522e+01  3.63423676e+03  1.00897667e+03  3.70584058e+02
-3.71563931e+03 -1.60942276e+03 -1.33320404e+04  1.85204759e+03]
```

```
#Fitting Ridge model for alpha = 20 and printing coefficients whi
alpha = 40
ridge = Ridge(alpha=alpha)

ridge.fit(X_train, y_train)
print(ridge.coef_)
```

```
[-1.31904431e+02  6.64870725e+01  4.29685726e-01  1.55419563e+03
-2.17867503e+03 -1.60578429e+03  0.00000000e+00  1.90795640e+03
 1.01215089e+04  5.52180239e+03  2.48896905e+02  1.51281698e+01
 2.63542436e+01  6.59757006e+03 -4.62583717e+03  2.17969494e+01
 3.64885878e+00 -1.33200306e+00  2.41138053e+01  2.63119317e+02
-1.57264792e+03  1.99859148e+01  2.95486461e+01 -1.45267345e+01
 3.50078264e+01  9.42668220e+02 -2.76660404e+03  8.47893460e+01
 1.34323870e+03 -5.01180056e+03 -5.32222919e+03  5.18666577e+03
 6.97382740e+02  4.66935416e+03  3.53407373e+03  7.08237125e+01
 4.78328183e+03  1.12912040e+01  1.63689556e+01  2.91653552e+01
-5.56837753e+00  2.34007971e+01  3.16111580e+01  9.80443797e+01
-1.60895513e+00  2.10010859e+02  2.88294887e+02  3.28523223e+03
 2.04742705e+03 -5.48472763e+02  2.51080838e+03 -1.50369763e+03
 1.44274220e+03 -1.25335887e+02  4.17904926e+03 -9.41623956e+02
-8.79619338e+02  4.99550399e+02  4.02916393e+02  5.90667391e+02
 4.85013899e+03 -2.74103995e+03 -3.83133245e+03  6.77742709e+03
-2.39895098e+03 -2.90409372e+03  1.11816926e+02  1.05370887e+01
-4.58236230e+03 -3.77691952e+03  1.91201349e+03 -6.09878922e+03
 5.45008302e+03  8.15492841e+03 -2.39907599e+03 -1.01746676e+03
 3.34524689e+02 -2.97182879e+03  1.39091224e+03  7.35092307e+03
-3.62720500e+03  1.39137248e+03 -4.36034867e+02  4.02066202e+03
 1.62050506e+02  4.03661849e+01 -2.90864649e+03  1.69947083e+03
-4.92577428e+02 -1.40298082e+02 -4.91928200e+02 -7.35429788e+02
 5.35284276e+02  0.00000000e+00  0.00000000e+00 -1.14181261e+02
 4.99614379e+02  8.84179732e+02 -3.55481135e+03 -1.99004556e+03
 1.48875409e+03  1.47669712e+03  1.39249948e+03  9.78433463e+02
-4.42627500e+02 -7.69572196e+02  1.75631961e+03  3.26404252e+03
-5.19913153e+02  7.34325348e+02  1.65131557e+03  3.18938978e+02
 1.04516936e+02  1.18664897e+03  0.00000000e+00  9.62600188e+01
-3.13118711e+02 -3.14758740e+03 -9.12265167e+02  3.09006229e+03
-4.66572547e+02  4.55944557e+02  5.23315029e+03 -3.94917384e+01
 2.39736699e+03 -1.83601358e+03 -1.93825423e+02  2.76594426e+02
-3.32451177e+03 -5.57410784e+02  3.69302710e+02 -3.68894951e+02
-2.93161794e+03  6.11548987e+02 -6.79669712e+02  1.36119553e+03
 1.03588703e+03 -3.94917384e+01  1.45430120e+03 -2.62526479e+03
 3.53248830e+03  3.16356815e+02 -1.36484578e+02 -4.40527121e+03
 7.56190736e+02 -7.31392905e+02 -2.17260705e+02  5.11926991e+02
-1.07245983e+03 -2.41240448e+03  3.39869045e+03  2.16692751e+03
-9.82500161e+02  3.00929930e+03  9.89695947e+02  1.56713666e+02
-2.20646555e+03 -2.23840402e+02 -1.07536418e+04  1.69663986e+03]
```

Impact on Coefficients –

Coefficients are **reduced** as we doubled the alphas value. If we increase it further, they get pushed to 0 thus may lead to underfitting

```
r2_score_train : 0.921229841583426
r2_score_test  : 0.6863152914107815
RSS_train      : 465516643892.1036
RSS_test       : 1002976660195.6322
MSE_train      : 459089392.3985242
MSE_test       : 2305693471.714097
```

```
r2_score_train : 0.9163469165400786
r2_score_test  : 0.6791698110925273
RSS_train      : 494373801529.57623
RSS_test       : 1025823645684.1133
MSE_train      : 487548127.74119943
MSE_test       : 2358215277.4347434
```

Impact on R2score and Residuals-

R2Score has decreased for train and test and the Residuals Sum of squares and mean for both test and train have increased, indicating more error

2. Lasso

Alpha=100	Alpha=200
<pre>alpha =100 lasso = Lasso(alpha=alpha) lasso.fit(X_train, y_train) print(lasso.coef_)</pre>	<pre>alpha =200 lasso = Lasso(alpha=alpha) lasso.fit(X_train, y_train) print(lasso.coef_)</pre>
<pre>[-1.23898174e+02 7.96429307e+01 4.05644629e-01 0.00000000e+00 -1.63584334e+03 -5.13193145e+02 0.00000000e+00 2.01885062e+03 9.47387349e+03 5.66064338e+03 2.50029392e+02 1.12372382e+01 1.99437753e+01 6.21678322e+03 -4.62372604e+03 3.30025310e+01 1.65559588e+01 1.15700790e+01 8.86837799e+00 -0.00000000e+00 -1.80323829e+02 4.60477960e+01 5.25034797e+01 1.44586514e+01 8.30637287e+00 5.98046596e+02 -2.85498726e+03 0.00000000e+00 5.76574001e+02 -4.54278133e+03 -9.31142707e+03 4.97612637e+03 1.02308562e+03 4.69459637e+03 4.07237763e+03 9.02313770e+01 4.90670359e+03 7.75380813e+00 1.44894680e+01 2.91052415e+01 -8.68155861e+00 2.29827894e+01 3.06057821e+01 1.02403346e+02 -1.95510024e+00 1.53586796e+02 3.88622355e+02 0.00000000e+00 0.00000000e+00 -0.00000000e+00 1.92123276e+03 -8.72407697e+02 0.00000000e+00 0.00000000e+00 6.76839510e+03 -0.00000000e+00 -0.00000000e+00 5.66817941e+02 0.00000000e+00 0.00000000e+00 8.49948112e+03 -0.00000000e+00 -0.00000000e+00 1.68706265e+04 -4.33661028e+02 -0.00000000e+00 0.00000000e+00 0.00000000e+00 -5.38176426e+03 -2.37001338e+03 0.00000000e+00 -5.85572949e+03 1.92171755e+04 1.91612783e+04 -3.22858067e+02 -0.00000000e+00 4.24153611e+02 -0.00000000e+00 7.78952383e+03 2.67271789e+04 -5.63479476e+02 0.00000000e+00 -0.00000000e+00 6.78750275e+03 0.00000000e+00 0.00000000e+00 -3.30672625e+03 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 -2.34745155e+03 -2.20967098e+03 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 -0.00000000e+00 0.00000000e+00 3.20733682e+03 -0.00000000e+00 0.00000000e+00 8.78165683e+02 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 -7.11880527e+03 -0.00000000e+00 1.35153650e+04 -0.00000000e+00 0.00000000e+00 1.10228118e+04 -0.00000000e+00 3.84443786e+03 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 -2.57836489e+03 0.00000000e+00 -0.00000000e+00 0.00000000e+00 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 -4.12667015e+03 7.92273805e+03 0.00000000e+00 -0.00000000e+00 -6.16049520e+03 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 -4.54198589e+02 4.00731443e+03 1.14028438e+03 0.00000000e+00 2.36788328e+03 0.00000000e+00 0.00000000e+00]</pre>	<pre>[-1.32942783e+02 6.13327755e+01 4.29752584e-01 0.00000000e+00 -1.60083245e+03 -0.00000000e+00 0.00000000e+00 0.00000000e+00 1.03144919e+04 5.21329398e+03 2.66713754e+02 1.67502606e+01 2.17996456e+01 7.41946162e+03 -4.74410929e+03 3.34979419e+01 1.22363738e+01 1.21355379e+01 9.01011176e+00 1.25250439e+02 -0.00000000e+00 4.74704964e+01 5.47766191e+01 1.58131425e+01 8.55747586e+00 2.62327360e+02 -1.76741301e+03 -0.00000000e+00 0.00000000e+00 -5.15947027e+03 -5.80699209e+03 4.98589269e+03 5.95856070e+02 4.67865324e+03 2.15585787e+03 8.82100311e+01 3.85027508e+03 1.17301870e+01 1.68834112e+01 3.02553710e+01 -8.95207121e+00 2.11197102e+01 3.20403935e+01 9.46436454e+01 -2.44658889e+00 1.45427601e+02 3.40637488e+02 0.00000000e+00 0.00000000e+00 -0.00000000e+00 0.00000000e+00 -1.40361035e+03 0.00000000e+00 0.00000000e+00 0.00000000e+00 3.61913674e+03 -0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 6.45926224e+03 -0.00000000e+00 -0.00000000e+00 1.38510967e+04 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 0.00000000e+00 -1.35795882e+03 -1.30487055e+03 0.00000000e+00 -4.57938314e+03 1.26354498e+04 1.41039050e+04 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 2.09812742e+03 1.83230814e+04 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 6.39903578e+03 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 -7.14557351e+01 -0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 3.87999959e+02 -0.00000000e+00 0.00000000e+00 1.11469577e+03 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 8.02471903e+02 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 0.00000000e+00 -0.00000000e+00 -6.49050100e+02 0.00000000e+00 -0.00000000e+00 0.00000000e+00 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 -3.02366624e+03 0.00000000e+00 0.00000000e+00 -0.00000000e+00 -5.73558904e+03 0.00000000e+00 -0.00000000e+00 -0.00000000e+00 -0.00000000e+00 -0.00000000e+00 -2.59507301e+03 2.43271703e+03 0.00000000e+00 -0.00000000e+00 2.44445250e+03 0.00000000e+00 0.00000000e+00]</pre>
Impact on Coefficients – Coefficeints are increasing and some are reducing as we doubled the alphas value.	
<pre>r2_score_train : 0.9212087318548419 r2_score_test : 0.7005309516932681 RSS_train : 465641398370.21454 RSS_test : 957523454852.1882 MSE_train : 459212424.42821944 MSE_train : 2201203344.487789</pre>	<pre>r2_score_train : 0.9141500574486914 r2_score_test : 0.690076457502834 RSS_train : 507356820630.77014 RSS_test : 990950693668.8806 MSE_train : 500351894.1131855 MSE_train : 2278047571.6525993</pre>
Impact on R2score and Residuals- R2Score has decreased for train and test and the Residuals Sum of squares and mean for both test and train have increased, indicating more error	

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

Both Ridge and Lasso are pushing the coefficients towards 0, regularizing it.

Since there are a lot of variables and Lasso has some coefficients 0. We chose Lasso in feature selection (135 lesser variables) and keeping the model simple. Its R2 score is also better than Ridge.

So, our choice of Model is lasso in this case.

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

Next 5 most important predictor variables will be-

- SaleCondition_Partial,
- BsmtExposure_Gd,
- RoofMatl_Tar&Grv,
- Neighborhood_Timber,
- Neighborhood_Crawfor

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

For a model to be robust and generalizable, it should be able to perform well on the unseen data.

A good R2 score on the test data will show that the model can predict even on unseen data.

There is a trade between bias and variance. Bias means how much error a model gives, and variance means sensitivity of data to input. Variance refers to changes in the model when trained on a different data set rather than the variance in the predicted values of a single model. If we try to decrease one the other increases and vice versa. So a good model is the one where bias and variance meet.

If a Model has less errors that means it has probably learnt the whole training set, that's why its performing so well. But that also means when the model is used on a test set, it will result in high variance as the data that model had learnt, has changed. This model will be complex.

Same way if a model has less variance that means that model is general and can perform good on unseen data, and the complexity of such model will be less.

