

Assignment-based Subjective Questions:

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?

The Optimal values of alpha for the ridge model I found is 20, and for the Lasso model it was 0.0001, when I doubled the values of the alpha in Ridge I noticed a difference on the coefficients and the scores became slightly better :

Before:

```
[ 0.08095241 0.16299651 0.08206225 0.16035367 0.05130647 0.0207292
-0.01438013 -0.06794143 0.15106186 0.32857089 0.03551852 -0.0578409
0.05377337 0.06729212 0.04309237 0.03827211 0.02187076 0.07165606
0.0141729 -0.0664515 -0.07565424 -0.05760499 -0.03170335 -0.05887295
-0.10644437 -0.06749259 -0.041513 -0.04990593 -0.04468605 0.04874868
0. -0.06772475 0. -0.04236721 -0.02952901 -0.05526044
-0.03813292 0.26538965 0. -0.05311207 0.06371278 0.18702262
0.10447821 0.14434546 -0.00063009 -0.06971114 0. -0.04970243
-0.06285498 -0.06331786 -0.10853127 -0.03083841 -0.00063009 0.05182329
-0.09231123 -0.1027105 0.01499389 0.02852825 0. -0.03191725
-0.08265652 -0.12565168 -0.03272755 0.0529219 0.0622162 0.00414889
0.03756009 0.00819052 0.0301166 -0.04462233 -0.00783034 -0.01700631
-0.05136881 0.00446337 -0.00236464 0.00289582 0.0139538 0.04042085
0.03366029 0.05399633]
```

```
-----Train Set-----
R2 Score : 0.9263237094406647
RSS Score : 75.22349266108137
MSE Score : 0.07367629055933532

-----Test Set-----
R2 Score : 0.8964698961441119
RSS Score : 45.346185488879016
MSE Score : 0.10353010385588816
```

After:

```
[ 8.18877380e-02 1.69099440e-01 7.75228526e-02 1.41499977e-01
5.44827424e-02 4.81226403e-02 -4.77217117e-03 -3.99964818e-02
1.26612598e-01 3.14583035e-01 3.32823171e-02 -5.57092468e-02
5.62874521e-02 7.29628284e-02 4.44341222e-02 3.05841959e-02
1.56832680e-02 5.20938010e-02 -3.80136021e-03 -6.30075061e-02
-7.37715935e-02 -5.51569856e-02 -2.83324378e-02 -5.61437814e-02
-1.02010573e-01 -6.43655415e-02 -4.11537910e-02 -4.80768616e-02
-4.22898629e-02 4.95924624e-02 0.00000000e+00 -6.46100971e-02
0.00000000e+00 -3.99620139e-02 -2.70873137e-02 -4.66605133e-02
-2.81109498e-02 1.81613760e-01 0.00000000e+00 3.36069543e-02
4.45999708e-02 1.29160053e-01 7.09318310e-02 1.05888002e-01
-1.45970650e-04 -6.10988937e-02 0.00000000e+00 -4.14224992e-02
-5.56179166e-02 -4.88684538e-02 -9.41905623e-02 -2.72302631e-02
-1.45970650e-04 4.48356870e-02 -8.46604974e-02 -9.69780794e-02
1.47456841e-02 2.77726323e-02 0.00000000e+00 -2.77765591e-02
-7.00952707e-02 -1.14379509e-01 -3.11854958e-02 5.08403105e-02
4.19796974e-02 -2.12754618e-04 3.11274459e-02 2.99600226e-03
8.79663137e-03 -3.08000717e-02 9.72182658e-05 -1.30292245e-02
-3.47992188e-02 -2.12209775e-03 -4.80891840e-03 3.06819880e-04
6.10519188e-03 4.10002121e-02 3.24077312e-02 5.34963490e-02]
```

```
-----Train Set-----
R2 Score : 0.9228117976944105
RSS Score : 78.80915455400688
MSE Score : 0.0771882023055895

-----Test Set-----
R2 Score : 0.9061724308207197
RSS Score : 41.09647530052475
MSE Score : 0.09382756917928026
```

Same goes for the Lasso:

Before:

```
[ 7.66699653e-02  1.51615667e-01  8.70507942e-02  1.84538180e-01
 4.66763406e-02 -1.63733491e-01 -7.97673066e-02 -2.58547286e-01
 3.28670949e-01  3.49432330e-01  3.92392993e-02 -5.85002031e-02
 5.21706182e-02  4.93649407e-02  5.14525983e-02  7.61571712e-02
 4.63958583e-02  1.52496793e-01  8.79999330e-02 -7.00737866e-02
 -7.51606838e-02 -5.77379994e-02 -3.77348112e-02 -6.08111962e-02
 -1.11985382e-01 -6.95094767e-02 -4.18523603e-02 -5.08939322e-02
 -4.71484637e-02  4.65431568e-02  0.00000000e+00 -6.82183426e-02
 0.00000000e+00 -4.45662420e-02 -3.26171746e-02 -6.06002290e-02
 -4.33910031e-02  4.66428560e-01  0.00000000e+00  1.01051755e-01
 1.09850116e-01  3.29649516e-01  1.86647893e-01  2.34914032e-01
 -2.94250633e-04 -8.50127441e-02  0.00000000e+00 -6.25495789e-02
 -7.63909951e-02 -8.71702825e-02 -1.29336193e-01 -3.66329225e-02
 -2.59336599e-03  5.96510484e-02 -9.91925797e-02 -1.08108631e-01
 1.44450086e-02  3.19204474e-02  0.00000000e+00 -3.75879515e-02
 -9.94002400e-02 -1.40915688e-01 -3.54173024e-02  5.66518214e-02
 1.89913436e-01  3.37966365e-02  9.77528956e-02  3.34579503e-02
 1.51707718e-01 -1.14642100e-01 -5.15727308e-02 -3.24362995e-02
 -1.37012019e-01  7.00829135e-02  3.22072230e-02  2.57382816e-02
 9.20879094e-02  3.87984150e-02  3.43200060e-02  5.50313501e-02]
```

```
-----Train Set-----
R2 Score : 0.9304594645764187
RSS Score : 71.00088666747655
MSE Score : 0.06954053542358134

-----Test Set-----
R2 Score : 0.8586605137401988
RSS Score : 61.90669498179291
MSE Score : 0.14133948625980117
```

After :

```
[ 7.72618600e-02  1.52818179e-01  8.71785837e-02  1.83634261e-01
 4.68176996e-02 -1.45622768e-01 -7.33277507e-02 -2.39829885e-01
 3.10926177e-01  3.49119078e-01  3.89327086e-02 -5.88780517e-02
 5.24232449e-02  5.03945949e-02  5.04018952e-02  7.01187355e-02
 4.26128684e-02  1.40023751e-01  7.67784033e-02 -6.93536536e-02
 -7.43853779e-02 -5.75266640e-02 -3.70307517e-02 -6.03590975e-02
 -1.11003100e-01 -6.93325371e-02 -4.10320483e-02 -5.03700044e-02
 -4.66312044e-02  4.63862069e-02  0.00000000e+00 -6.85326131e-02
 0.00000000e+00 -4.41857097e-02 -3.19736644e-02 -5.85972508e-02
 -4.17367002e-02  4.61289141e-01  0.00000000e+00  9.99454721e-02
 1.08724353e-01  3.26317642e-01  1.84698605e-01  2.34456029e-01
 -3.15930444e-04 -8.35193679e-02  0.00000000e+00 -6.14803073e-02
 -7.53318159e-02 -8.52445440e-02 -1.27202548e-01 -3.63186068e-02
 -2.39273471e-03  5.88235067e-02 -9.87442877e-02 -1.07713011e-01
 1.45911161e-02  3.15574185e-02  0.00000000e+00 -3.75125251e-02
 -9.98479367e-02 -1.41461863e-01 -3.51968756e-02  5.64733261e-02
 1.69274086e-01  2.87012174e-02  8.71944123e-02  2.94867852e-02
 1.32857660e-01 -8.94953739e-02 -3.58751858e-02 -2.74762216e-02
 -1.06555264e-01  4.29086257e-02  1.70448050e-02  1.60500387e-02
 5.94512170e-02  3.87694866e-02  3.41043170e-02  5.51279032e-02]
```

```
-----Train Set-----
R2 Score : 0.9303544424782978
RSS Score : 71.10811422965799
MSE Score : 0.06964555752170223

-----Test Set-----
R2 Score : 0.8608717685780226
RSS Score : 60.93816536282612
MSE Score : 0.13912823142197744
```

For the most important predictor variables after the change we have :

Ridge :

	Features	coefficients
9	GrLivArea	0.315
37	RoofMatl_CompShg	0.182
1	OverallQual	0.169
3	YearBuilt	0.141
41	RoofMatl_Tar&Grv	0.129
8	TotalBsmtSF	0.127
43	RoofMatl_WdShngl	0.106
0	LotArea	0.082
2	OverallCond	0.078
13	GarageCars	0.073

Lasso:

	Features	coefficients
37	RoofMatl_CompShg	0.461
9	GrLivArea	0.349
41	RoofMatl_Tar&Grv	0.326
8	TotalBsmtSF	0.311
43	RoofMatl_WdShngl	0.234
42	RoofMatl_WdShake	0.185
3	YearBuilt	0.184
64	GarageType_Attchd	0.169
1	OverallQual	0.153
17	MSZoning_RL	0.140

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?

I will choose to apply the doubled value since it increases my R2 Score in the test set for both the Lasso and Ridge as showing above.

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?

Removing the 5 important predictors from the Lasso model :

	Features	coefficients
37	RoofMatl_CompShg	0.461
9	GrLivArea	0.349
41	RoofMatl_Tar&Grv	0.326
8	TotalBsmtSF	0.311
43	RoofMatl_WdShngl	0.234

The R2 Score dropped on the train dataset :

```
-----Train Set-----  
R2 Score : 0.8783060044834927  
RSS Score : 124.24956942235391  
MSE Score : 0.12169399551650727  
  
-----Test Set-----  
R2 Score : 0.8592481548953999  
RSS Score : 61.649308155814865  
MSE Score : 0.14075184510460015
```

And we got the new 5 important predictors :

	Features	coefficients
1	OverallQual	0.294
5	BsmtFinSF1	0.194
0	LotArea	0.156
71	GarageCond_TA	0.120
15	MSZoning_RL	0.118

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?

Making the model robust and generalizing well is to make it simple not too simpler, to avoid neither Overfitting or Underfitting problems, and use regularization to control the trade-off between the variance and bias to end up with more generalizing model and get good accuracy scores that it's our reference to be in a good balance of this trade-off.