

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?

Ans:

alpha for ridge = 1.0

alpha for ridge = 0.0001

changes after alpha doubled in ridge

```
Train Set:
r2 score: 0.9159793056621017
nMSE score: 0.0027466089483170067
MAE score: 0.03754050700914243
RMSE score: 0.05240810002582622
Test Set:
r2 score: 0.8693689325530792
nMSE score: 0.004289226931103931
MAE score: 0.0500345962440427
RMSE score: 0.06549218984813328
```

changes after alpha doubled in lasso

```
Train Set:
r2 score: 0.9155105695483937
nMSE score: 0.002761931778179996
MAE score: 0.037640271578771464
RMSE score: 0.052554084314922624
Test Set:
r2 score: 0.864350245593547
nMSE score: 0.00445401382051944
MAE score: 0.05103185476079498
RMSE score: 0.06673839839642123
```

Top 10 params for ridge

Features	Coefficient	Abs_Coefficient_Ridge
GrLivArea	0.3217	0.3217
OverallQual	0.1932	0.1932
Neighborhood_StoneBr	0.1040	0.1040
GarageCars	0.0993	0.0993
OverallCond	0.0967	0.0967
TotalBsmtSF	0.0879	0.0879
Foundation_Slab	0.0851	0.0851
BsmtFinSF1	0.0791	0.0791
Neighborhood_NridgHt	0.0758	0.0758
MasVnrArea	0.0688	0.0688

Top 10 params for lasso

Features	Coefficient	Abs_Coefficient_Ridge
GrLivArea	0.3778	0.3217
OverallQual	0.2166	0.1932
Neighborhood_StoneBr	0.1079	0.1040
GarageCars	0.0976	0.0993
OverallCond	0.1017	0.0967
TotalBsmtSF	0.0871	0.0879
Foundation_Slab	0.0782	0.0851
BsmtFinSF1	0.0733	0.0791
Neighborhood_NridgHt	0.0861	0.0758
MasVnrArea	0.0624	0.0688

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?

Ans:

alpha for ridge = 1.0
alpha for ridge = 0.0001

- The r^2 square for the ridge and lasso regression model is almost the same.
- The MSE score of test data set slightly higher for lasso regression. So lasso will work better than ridge regression for unseen data

Thus Lasso regression has works slightly better than ridge

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?

Top 5 params for lasso			
	Features	Coefficient	Abs_Coefficient_Ridge
	TotalBsmtSF	0.2171	0.2171
	FullBath	0.1390	0.1390
5	Foundation_Slab	0.1123	0.1123
	Neighborhood_MeadowV	-0.1115	0.1115
	dummy_ExterQual	0.1094	0.1094

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?

Ans:

Robust models are simple models at the same time they will be robust. It is a balance between bias and variance. When the complexity of model increases, it's variance also increases. But at the same time it's error will decrease. Vice versa simple models error will be high and variance will be low. A good model should balance between this bias and variance and that model will be robust.

Bias-Variance Tradeoff

