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:

Regularization equation

$$\sum_{i=1}^{n}(y_i - \beta(x_i))^2 + \lambda\sum_{i=1}^{n}\beta_i$$

$\sum_{i=1}^{n}(y_i - \beta(x_i))^2$ $\ error\ term$ $\ \lambda\sum_{i=1}^{n}\beta_i$ $\ regularization\ term$

λ presents alpha or hyperparameter

Optimal value of alpha for ridge and lasso regression is the value of hyper parameter at which model predictions on both train and test data is good. It helps in creation of a good model without over or under-fit.

In assignment optimal value of alpha for ridge is 100 and for lasso is 0.01

When we double the value of alpha of lasso and ridge model, it penalizes the coefficient more i.e. makes the coefficient tends to zero in ridge and in lasso it makes coefficients 0

As λ increase weight of regularization term increase which tends to make model generalized

This results in high bias and low variance and sometimes results in under-fit.

Ridge most important predictors before doubling alpha

	Coefficients	Abs_Coefficients
OverallQual	0.085185	0.0852
GrLivArea	0.073105	0.0731
1stFlrSF	0.058196	0.0582
Neighborhood_NridgHt	0.055543	0.0555
TotalBsmtSF	0.054262	0.0543
KitchenQual	0.053557	0.0536
TotRmsAbvGrd	0.053588	0.0536
Neighborhood_NoRidge	0.053157	0.0532
GarageCars	0.051224	0.0512
ExterQual	0.048624	0.0486
[illegible]	[illegible]	[illegible]

Ridge most important predictors after doubling alpha

	Coefficients	Abs_Coefficients
OverallQual	0.069001	0.0690
GrLivArea	0.062568	0.0626
1stFlrSF	0.051077	0.0511
KitchenQual	0.048175	0.0482
TotalBsmtSF	0.047588	0.0476
TotRmsAbvGrd	0.046070	0.0461
ExterQual	0.045263	0.0453
GarageCars	0.044192	0.0442
Neighborhood_NridgHt	0.041952	0.0420
Neighborhood_NoRidge	0.041056	0.0411

Number of features having value close to zero increased in Ridge

In Lasso,

lasso most important predictors before doubling alpha

	Coefficients	Abs_Coefficients
GrLivArea	0.2476	0.2476
OverallQual	0.1918	0.1918
Neighborhood_NridgHt	0.1198	0.1198
GarageCars	0.0971	0.0971
Neighborhood_NoRidge	0.0868	0.0868
Neighborhood_StoneBr	0.0832	0.0832
BsmtFinSF1	0.0741	0.0741
KitchenQual	0.0597	0.0597
BsmtExposure	0.0567	0.0567
TotRmsAbvGrd	0.0539	0.0539

lasso most important predictors after doubling alpha

	Coefficients	Abs_Coefficients
GrLivArea	0.2553	0.2553
OverallQual	0.2072	0.2072
Neighborhood_NridgHt	0.1028	0.1028
GarageCars	0.0946	0.0946
BsmtFinSF1	0.0839	0.0839
Neighborhood_NoRidge	0.0793	0.0793
Neighborhood_StoneBr	0.0674	0.0674
KitchenQual	0.0674	0.0674
ExterQual	0.0583	0.0583
BsmtExposure	0.0484	0.0484

```
lasso_coefficients_df[lasso_coefficients_df['Coefficients']!=0].sort_values(by='Coefficients',ascending=False).shape
```

```
(93, 2)
```

```
lasso_coefficients_df_02[lasso_coefficients_df_02['Coefficients']!=0].sort_values(by='Coefficients',ascending=False).shape
```

```
(60, 2)
```

Before doubling number of non-zero features are 93 but after doubling it decreased to 60

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:

In assignment optimal value of lambda is choosed by running grid search with various lambda values ranging from 0.001 to 2000. Based on mean train and test score we choose lambda values.

After determining optimal value of lambda

I choosed lasso regression, because lasso regression along with regularization it also acts as a feature selectior, which reduces the number of features by making the coefficients of features to 0.

Model selection metrics are good for lasso compare to ridge

Metrics	**Lasso**	**Ridge**
R2score – train	0.876	0.855
R2 score – test	0.879	0.853
Mean error square train	0.124	0.145
Mean error square test	0.138	0.168

mean error score- R2 score is higher and mean square error is lower than ridge. Lasso gives better predications than ridge and increase performance as it gives sparse coefficients matrix.

Question 3

After building the model, you realized 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:

After rebuilding the model by removing the 5 most important variables.

	Coefficients	Abs_Coefficients
2ndFlrSF	0.2457	0.2457
1stFlrSF	0.1840	0.1840
BsmtQual	0.0959	0.0959
KitchenQual	0.0940	0.0940
TotalBsmtSF	0.0939	0.0939

Question 4

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

Answer:

Using regularization, we can make model robust and generalizable by tuning the hyperparameter.

Hyperparameter restricts model to not become overfitted by learning all the data points. Because if a model is overfitted it performs well on training data but gives poor results on testing data or unseen data. So, it can't be robust.

Model should be simple, and it should have trade off between bias and variance. It should be a good model with under or over fit.

In validation we should use cross validation, as if we perform validation on testing data by building several models model may try to seek peek the pattern in test data and becomes not generalized.

By making model robust and generalizable, it performs well on the unseen data.

As we are using regularization, its accuracy decreases on the training data compare to a overfitted model but it's accuracy on unseen high than overfitted model.

So, its important in regularization to tune the hyperparameter to get a value which makes model to have good accuracy and robust.

Note: As hyperparameter increase error term increases i.e. bias increase and variance decrease.

So, we should choose a hyperparameter with good bias and variance as they constitute to the stability and generalization of model.