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?

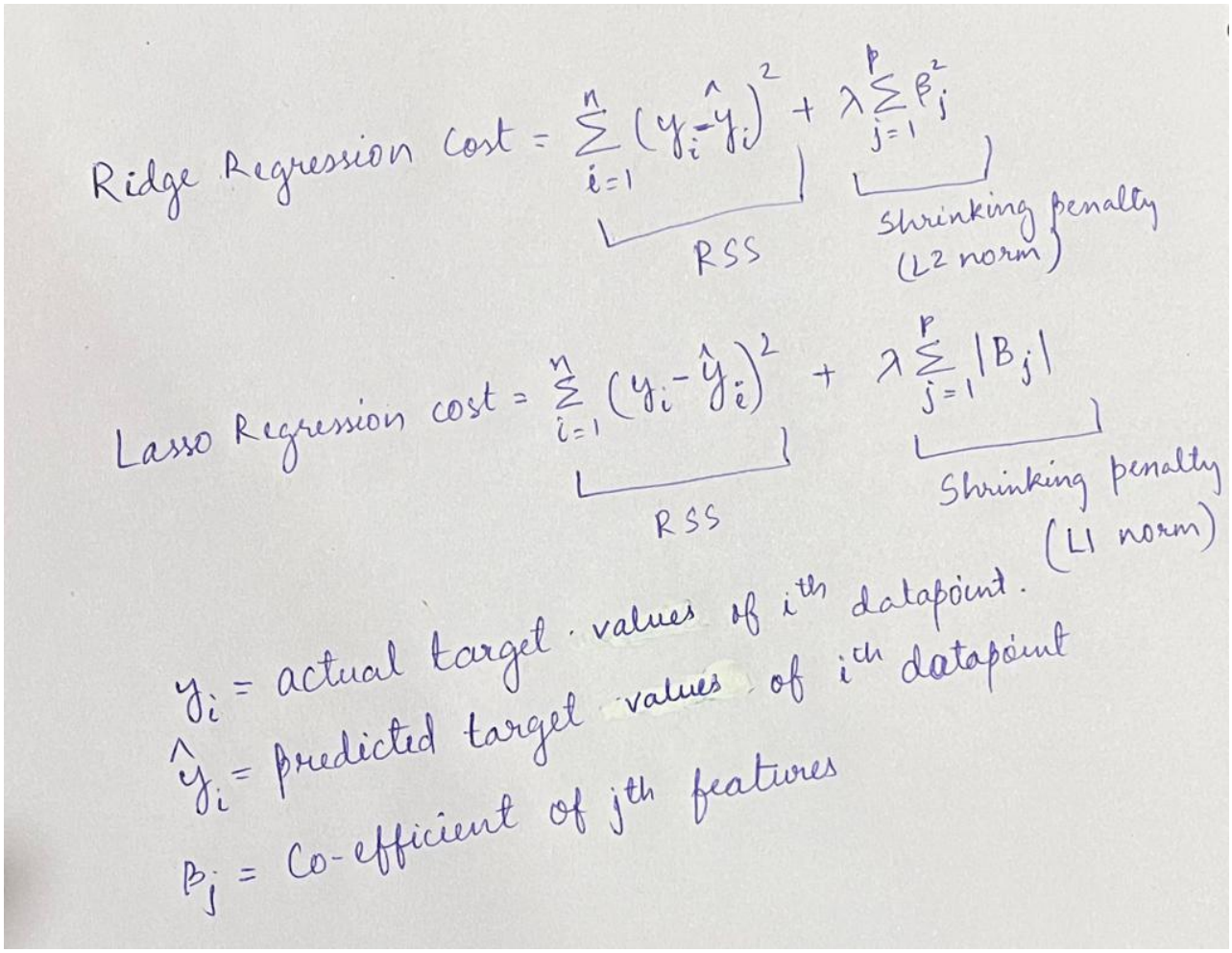
Answer:

Optimal value of alpha:

- Optimal alpha (lambda) value for Lasso Regression model is: .0007
- Optimal alpha (lambda) value for Ridge Regression model is: 8

Effect of choosing double the value of optimal alpha:

Formula for cost functions of Ridge and Lasso:



It is important to consistently set the parameters and improve the prediction accuracy even if the variance decreases, and to keep the model interpretable

Ridge regression, uses a tuning parameter called lambda because the penalty is the square of the magnitude of the coefficients determined by cross validation. The remaining sum or squares can be reduced by using a stick. The penalty is the sum of the squares of the coefficients beyond the lambda, so those expensive coefficients are penalized. As we increase the value of lambda, the variance of the model decreases and the bias remains constant. Ridge regression, unlike Lasso Regression, includes all variables in the final model.

Lasso regression, uses the tuning parameter lambda as the penalty is the maximum of the absolute value of the coefficients determined by cross validation As the value of lambda increases, the Lasso shrinks the coefficient to zero and equalizes the variables to exactly 0. Lasso is also a variable selection option. When the lambda value is small, it is a simple linear regression and as the lambda value increases, shrinkage occurs and variables with values of 0 are ignored by the model.

Most important predictor variables after the change is implemented:

Top 10 features with beta coefficient values obtained from Ridge after using alpha= 16

```
OverallQual             0.193234
GrLivArea               0.157812
1stFlrSF                0.119808
GarageArea              0.105336
2ndFlrSF                0.103908
OverallCond             0.102095
FullBath                0.092758
Neighborhood_Edwards   -0.081767
MSSubClass_30          -0.081295
YearRemodAdd            0.077295
dtype: float64
```

Top 10 features with beta coefficient values obtained from Lasso after using alpha= .0012.

```
GrLivArea               0.414665
OverallQual             0.338961
GarageArea              0.140679
OverallCond             0.117321
Neighborhood_Somerst    0.103411
Neighborhood_Crawfor    0.094462
Neighborhood_NridgHt    0.089895
MSSubClass_30          -0.081067
YearBuilt               0.080022
YearRemodAdd            0.071770
dtype: float64
```

So, after Doubling value of alpha the most important variable:
In Ridge model: OverallQual (Rates the overall material and finish of the house)
In Lasso model: GrLivArea (Above grade (ground) living area square feet)

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: A model should not be unnecessary complex. Model complexity depends on two main things: No. of features or independent variables and Magnitude of beta coefficients. Normalization (Ridge and Lasso) already shrinks beta coefficients towards zero. Now, Lasso and Ridge both have similar r2 score and MAE on test dataset. But Lasso has eliminated 123 features and final no. of features in Lasso Regression model is 103. Where Ridge has all 226 features. So, the Lasso model is simpler than Ridge with having similar r2 score and MAE.

Ridge:

```
r2 score on testing dataset: 0.8820724114129794
MSE on testing dataset: 0.019436209067663194
RMSE on testing dataset: 0.13941380515452262
MAE on testing dataset: 0.09335836776845349
```

Lasso:

```
r2 score on testing dataset: 0.8811474224048781
MSE on testing dataset: 0.019588660923604284
RMSE on testing dataset: 0.13995949743981037
MAE on testing dataset: 0.09315263954899028
```

As these two models shows almost similar performance on test dataset, we should choose the simpler model. So, I will choose Lasso as my final model.

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- Initially top 5 features in Lasso model are as below:

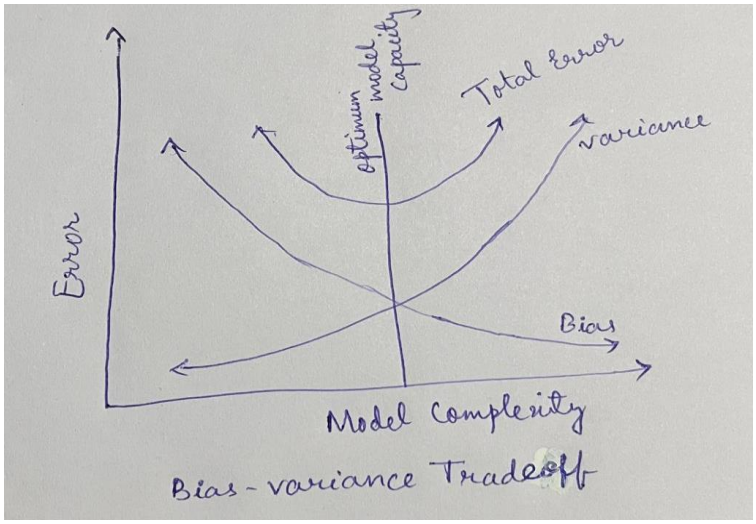
```
GrLivArea              0.441183
OverallQual            0.299657
GarageArea             0.135930
OverallCond            0.124396
Neighborhood_Somerst   0.113196
dtype: float64
```

After dropping GrLivArea, OverallQual, OverallCond, GarageArea, Neighborhood_Somerst features, rebuilt the Las so model again with rest of the features, now 5 most important predictor variables are as below.

```
1stFlrSF                    0.422404
2ndFlrSF                    0.338779
BsmtExposure_Not Present   -0.134563
YearRemodAdd                0.127841
MSSubClass_30              -0.123846
dtype: float64
```

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:
Bias- Variance Tradeoff:



A model must be robust enough to detect data patterns in the training data set but it is also not. The difficulty is that it also learns words on the training data set. The model must be sufficiently generalized and it is so simple that it memorizes all the data points in the training data set.
Under fitting models tend to have higher bias and lower variance. It cannot find data structures training dataset, so it does not perform badly on both training and testing datasets. whereas the overfitting model.
In general, there is low bias and high variance. It performs well on the training dataset but not badly in testing dataset or unseen data.
The overfitting condition can be easily identified by comparing model performance in training and testing data set. If there is a significant difference in model performance (r2 score, model accuracy, MAE, RMSE, Confusion Matrix and other analytical metrics) training and testing dataset then it is a matter of eligibility Over qualification.
A robust model should have low bias and low variance with its under fitting and eligibility over qualification. It can be obtained by trading off bias and variance. One way to do is to develop a robust and generalizable model, model complexity must be reduced to overcome overfitting.
Model complexity depends on two main factors: the number of features or the number of independent variables and the size of beta coefficients. Normalization (Ridge and Lasso) already reduces the beta coefficient.
It's going to zero. Additionally, Lasso also helps reduce volumes by reducing certain beta coefficients to exact 0. thus helping to prevent overfitting. A robust and generalizable model must have accuracy. We are almost identical/close in all types of training and testing data.