**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

Optimum values of Ridge and Lasso Regression is 4,0.0001 respectively.

Present top feature are

RIDGE

| | |
|---|---|
| GrLivArea | 0.097719 |
| OverallQual | 0.082223 |
| 1stFlrSF | 0.074475 |
| 2ndFlrSF | 0.072556 |
| TotalBsmtSF | 0.070311 |
| GarageArea | 0.060763 |
| Neighborhood_StoneBr | 0.046350 |
| BsmtFinSF1 | 0.044967 |
| OverallCond | 0.042563 |

LASSO

| | |
|---|---|
| GrLivArea | 0.263832 |
| OverallQual | 0.118785 |
| TotalBsmtSF | 0.079547 |

```
OverallCond            0.062059
GarageArea             0.060684
YearBuilt              0.056170
Neighborhood_StoneBr    0.052078
Neighborhood_Crawfor    0.040101
Neighborhood_NoRidge    0.037455
```

After the changes mentioned above are implemented still top predictor values will not be affected as changing alpha doesn’t affect the relative importance of predictor.If we increase alpha model becomes simpler as more variables have coefficients near to 0 in both Ridge and lasso. So too much increase in alpha may cause model to loosed predicting power

**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

I will choose Lasso Regression for the following reasons

1)Both Ridge and Lasso Regressions have similar accuracy and error metrics but Lasso slightly outperforms Ridge in R2 score

2) Both Models have comparable test and train R2 scores which implies none of them are overfitting

3)Lasso Regression has eliminated more features and model is simpler than Ridge which clearly gives us reason to select Lasso as better fit in our presenting scenario

**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?

Ans

If five most important variables are absent.Re running model by removing those columns gives following five important predictors

Before

| | |
|---|---|
| GrLivArea | 0.263832 |
| OverallQual | 0.118785 |
| TotalBsmtSF | 0.079547 |
| OverallCond | 0.062059 |
| GarageArea | 0.060684 |

After

```
1stFlrSF                0.276539

2ndFlrSF                0.166176

BsmtFinSF1              0.089468

Neighborhood_StoneBr    0.064855
```

```
BsmtUnfSF               0.062603
```

## 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

1)If we give too much importance to accuracy of model especially in case of training, this may result in model overfitting data

2)An overfit model learns everything about train data and looses its generalizability in predicting different data sets

3)As model overfits and complexity increases its variance also increases and bias reduces. This doesn't mean the model has all required properties to be a good model.Any variation in dataset results in vastly different metrics.

4)So A model which has decent accuracy but simple can perform better on test data as model doesn’t over fit and variance is low. Even though bias is high, Its still generalizable with predefined expectation and metrics