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 for ridge: 9

Optimal value for lasso: 100

On doubling the values

For Ridge: Coeff values are increasing as alpha is increasing

For Lasso: As alpha increase more features get removed from the model

Top Features after change is implemented:

With Ridge

```
Neighborhood_NoRidge    37768.949841
Neighborhood_NridgHt    36873.855166
OverallQual             35081.211855
1stFlrSF                29336.126842
2ndFlrSF                18854.677568
```

With Lasso

```
Neighborhood_NoRidge    56501.687327
Neighborhood_NridgHt    49362.166891
OverallQual             33809.798162
1stFlrSF                28535.340496
Neighborhood_Somerst    23049.767933
```

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:

Lasso as it gives feature selection option also. It removes unwanted features from model without affecting the model accuracy which makes the model generalized.

If we have too many variables and one of our primary goal is feature selection, then we will use Lasso.

If we don't want to get too large coefficients and reduction of coefficient magnitude is one of our prime goals, then we will use Ridge Regression.

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?

Answer:

Top 5 features are

```
Neighborhood_NoRidge    59938.254528
Neighborhood_NridgHt    54028.183545
OverallQual             32530.977162
Neighborhood_Veenker    29393.829646
1stFlrSF                28417.865743
```

After dropping next top 5 features after dropping 5 main predictors 1stFlrSF, MSSubClass_90, MSSubClass_120, TotalBsmtSF, HouseStyle_1Story

```
'Neighborhood_Somerst', 'Neighborhood_StoneBr', Exterior2nd_Wd Shng
, Exterior2nd_ImStucc', 'Exterior2nd_Stucco
```

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?

Answer:

- A model is robust when any variation in the data does not affect its performance much.
- A generalizable model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- To make sure a model is robust and generalizable, we have to take care it doesn't overfit. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.
- In other words, the model should not be too complex in order to be robust and generalizable.

- If we look at it from the perspective of Accuracy, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.
- In general, we have to find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso