## 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**:

The optimal value of alpha for both ridge and lasso regression is 0.0001. If we increase the alpha (hyper parameter value), the accuracy of the model starts dropping gradually. It might increase a bit till the optimal hyper parameter value but the accuracy will decrease with the increase in alpha and model will become more parse.

Features that lasso provides at optimal alpha(0.0001):

| Featuere | Coef |
|---|---|
| MSSubClass_160 | 1.661479 |
| Total_sqr_footage | 0.801819 |
| MSSubClass_70 | 0.729687 |
| EnclosedPorch | 0.464683 |
| PoolArea | 0.458387 |
| 3SsnPorch | 0.449309 |
| ScreenPorch | 0.424282 |
| LotConfig_FR2 | 0.416400 |
| LotFrontage | 0.382299 |
| MiscVal | 0.341944 |

Features that lasso provides at optimal alpha(0.0002):

| Featuere | Coef |
|---|---|
| MSSubClass_160 | 1.633332 |
| Total_sqr_footage | 0.704254 |
| MSSubClass_70 | 0.462230 |
| EnclosedPorch | 0.460270 |
| PoolArea | 0.452888 |
| 3SsnPorch | 0.449200 |
| ScreenPorch | 0.421660 |
| LotFrontage | 0.390213 |
| LotConfig_FR2 | 0.315798 |
| LotConfig_Inside | 0.307388 |

Features that ridge provides at optimal alpha(0.0001):

| Feaure | Coef |
|---|---|
| MSSubClass_160 | 1.688584 |
| Total_sqr_footage | 0.898970 |
| LotConfig_FR2 | 0.521824 |
| MiscVal | 0.469450 |
| EnclosedPorch | 0.468222 |
| PoolArea | 0.464391 |
| 3SsnPorch | 0.450001 |
| ScreenPorch | 0.427369 |
| LotFrontage | 0.402402 |
| Total_Bathrooms | 0.396804 |

Features that ridge provides at optimal alpha(0.0002):

| Feaure | Coef |
|---|---|
| MSSubClass_160 | 1.688553 |
| Total_sqr_footage | 0.898878 |
| LotConfig_FR2 | 0.521768 |
| MiscVal | 0.469423 |
| EnclosedPorch | 0.468224 |
| PoolArea | 0.464396 |
| 3SsnPorch | 0.450013 |
| ScreenPorch | 0.427380 |
| LotFrontage | 0.402384 |
| Total_Bathrooms | 0.396763 |

## 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:**

After creating model in both Ridge and Lasso we can see that the r2_scores are almost the same for both. However, since Lasso will penalize more on the dataset and can also helpo in feature elimination, therefore we are going to consider Lasso as our final model.

## 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:**

The five most important predictor variables in the Lasso model are below:

- MSSubClass_160 - 2-STORY PUD - 1946 & NEWER type of dwelling involved in the sale
- Total_sqr_footage - Overall area for all floors and basement
- MSSubClass_70 - 2-STORY 1945 & OLDER type of dwelling involved in the sale

- EnclosedPorch - Enclosed porch area in square feet
- PoolArea - Pool area in square feet

If above variables are not available in the incoming data, then the five most important variables will be:

- 3SsnPorch - Three season porch area in square feet
- ScreenPorch - Screen porch area in square feet
- LotConfig_FR2 - Frontage on 3 sides of property
- LotFrontage - Linear feet of street connected to property
- MiscVal - $Value of miscellaneous feature

## 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:**

We should make sure that the model is robust and generalisable by regularizing the model and using a regularisation term with RSS because the hyper parameter will ensure to srike the right balance between the model being too simple or too complex. Making the model more generalisable may have a negative impact on the accuracy upto some extent but we can also have a look at the precision and recall of the model because sensitivity and specificity also play imporant roles in the model evaluation criteria. Together if all three are above average we may accept the model. Also, a very accurate model may have a chance of getting overfitted.