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?**

Optimal value of alpha for ridge – 2

Optimal value of alpha for lasso **-** 0.00005

Optimal value of alpha can be found with the help of K-fold validation technique. Basically the idea is to use different sets of train and test sets and create models out of them. Like let's say we've dataset of 1000 rows with 10 attributes and we're using K-Fold validation of 10 and alpha (5, 10, 15, 20, 25, and 30) then we'll create 60 models and then compare values of RSME or any other predictor to determine value of alpha.

More the value of alpha better are the chances of overfitting the data (values of coefficients will increase which will lead to overfitting) so if you double the alpha then you'll increase the bias of the model.

We can use Negative RMSE to check for efficiency of the models after doing changes.

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?**

Both the regression models have their own benefits.

Basic idea behind ridge regression is by using optimal value of lambda you will get coefficient values of all the features which are there in model. So we'll use Ridge regression model when every feature in the model has same importance.

In Lasso regression it might be possible that some of the features have 0 coefficients which eventually telling that these features (have coefficients 0) are not significant in determining the target variable. So we'll lasso when we don't need all the features but only important ones.

So if I have too many attributes and not all are important then I'll choose Lasso and if I want all the attributes in the model then I'll use Ridge.

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?**

Five important Features – Total Area, OverAllQual, Lot Area, TotalBsmpSF, Age_of_house

Next 5 imp features - SaleCondition_Partial, MSZoning_RM, MSZoning_RH, BldgType_Duplex, and Heating_Grav

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?

Robustness of the model is depend on bias-variance trade off. Higher the bias lesser the variance and higher the variance of the model lesser the bias. So finding perfect tradeoff between these two things is most important.

In order find perfect trade off we can make use of Regularization techniques like below,

1) Ridge regression
2) Lasso regression

All these techniques add penalty in the error term (RSS) if you try to increase the number of features and eventually increase the complexity.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2 = \mathrm{RSS} + \lambda\sum_{j=1}^{p}\beta_j^2$$

In ridge regression cost function is modified by adding above penalty. In ideal model cost function needs to be minimized to get optimum values of the coefficients. So in Ridge regression we've to minimize the above equation. Lambda is the deciding factor of which decides overall value of coefficients (higher or lower). Higher the value of lambda greater the value of coefficients which will lead to overfitting. Optimum value of lambda can be found using K-fold validation.

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|.$$

Above is the equation for Lasso regression. We can see that only difference is in added penalty. Lasso regression decrease the value of some of the coefficients to 0