Question1. 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 -
GridSearchCV gives below lambda/alpha values as best hyperparameter values -

Ridge - 0.01 => Train r2 score - 0.946788, Test r2 score - 0.836535
Lasso - 0.0001 => Train r2 score - 0.904407, Test r2 score - 0.860556
Based on coefficients' value in both model, below are the important predictor variables -
Ridge - ('RoofMatl_Membran', 'RoofMatl_WdShngl', 'RoofMatl_Metal', 'RoofMatl_CompShg', 'RoofMatl_Tar&Grv')

Lasso - ('GrLivArea', 'RoofMatl_WdShngl', 'OverallQual', 'Neighborhood_NoRidge', 'GarageCars')

Now, if we double these lambda values, let's see impact on train and test r2 scores -
1. Ridge - 0.02 => Train - 0.945053 , Test - 0.841634
2. Lasso - 0.0002 => Train - 0.890517, Test - 0.863395

Based on coefficients' value in both model, below are the important predictor variables
1. Ridge
('RoofMatl_WdShngl', 'RoofMatl_Membran', 'RoofMatl_CompShg', 'RoofMatl_Metal', 'RoofMatl_Tar&Grv')
2. Lasso
('GrLivArea', 'OverallQual', 'RoofMatl_WdShngl', 'Neighborhood_NoRidge', 'GarageCars')

But when we see the best model ourselves (based on difference between train and test r2 scores) -
1. Ridge - 2.0 => Train - 0.902400, Test - 0.868895 (Calculating train test scores in next command for this lambda value)
2. Lasso - 0.0001 => Train - 0.904407, Test - 0.860556 (Same as above)

Important predictors for this Ridge model (which are almost similar to Lasso model) -
('GrLivArea', 'OverallQual', 'RoofMatl_WdShngl', '2ndFlrSF', 'Neighborhood_NoRidge')
Now, we can clearly see that this Ridge model (with parameter alpha = 2 is most similar to lasso one)

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 -
K-fold cross validation (GridSearchCV) has given best hyperparameters for Ridge and Lasso regression as 0.01 and 0.0001 respectively. The test r2 scores for Ridge and Lasso are 0.8365 and 0.8605 respectively. So, in that sense Lasso is doing better, so we will choose this. Also, we know that Lasso regression does feature selection and eliminates features which are not required. We have already seen in above commands that Lasso with alpha = 0.0001 has selected 61 columns and made all other features as 0. So, we will choose Lasso regression with alpha = 0.0001.

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 -
From answer 1, we found the 5 most important predictor variables (given by Lasso Regression) are -
('GrLivArea', 'OverallQual', 'RoofMatl_WdShngl', 'Neighborhood_NoRidge', 'GarageCars')

Now, let's try to drop these variables, train the lasso regression and again find the 5 most important predictor variables..
From below few commands, the alpha/lambda for lasso regression is still 0.0001 and results for same are –
lambda = 0.0001 , train r2 score = 0.8759 , test r2 score = 0.8299
And 5 most important predictor variables now are –
('TotalBsmtSF', '2ndFlrSF', 'LotArea', 'MasVnrArea', 'FullBath')

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 can make sure that model is robust and generalisable by taking care of bias – variance trade–off. The model should have balance between bias and variance, it should not be like – model is having very less training error (low bias) but has very high test error (high variance). We can have training accuracy a bit low (than overfit case) but test accuracy similar to training one. Here in our case (from 1st question), we have chosen ridge regression with lambda/alpha = 2, because for that case r2 score difference between train and test data is less as compared to ridge regression with lambda/alpha = 0.01. We can see that there are some implications on training score, because to avoid overfitting, our training score/accuracy will get lower a bit. But then our model will be more robust and works fine for unseen test data.