# Assignment Part-II

The following questions are the second part of the graded assignment. Please submit the answers in one PDF file. For writing normal text, please use MS Word (or similar software that can convert documents to PDF). For equations and figures, you can write/draw them on a blank sheet of paper using a pen, click images and upload them in the same Word document.

The final submission will be in the form of one PDF file. A sample PDF to illustrate the submission format is provided below.

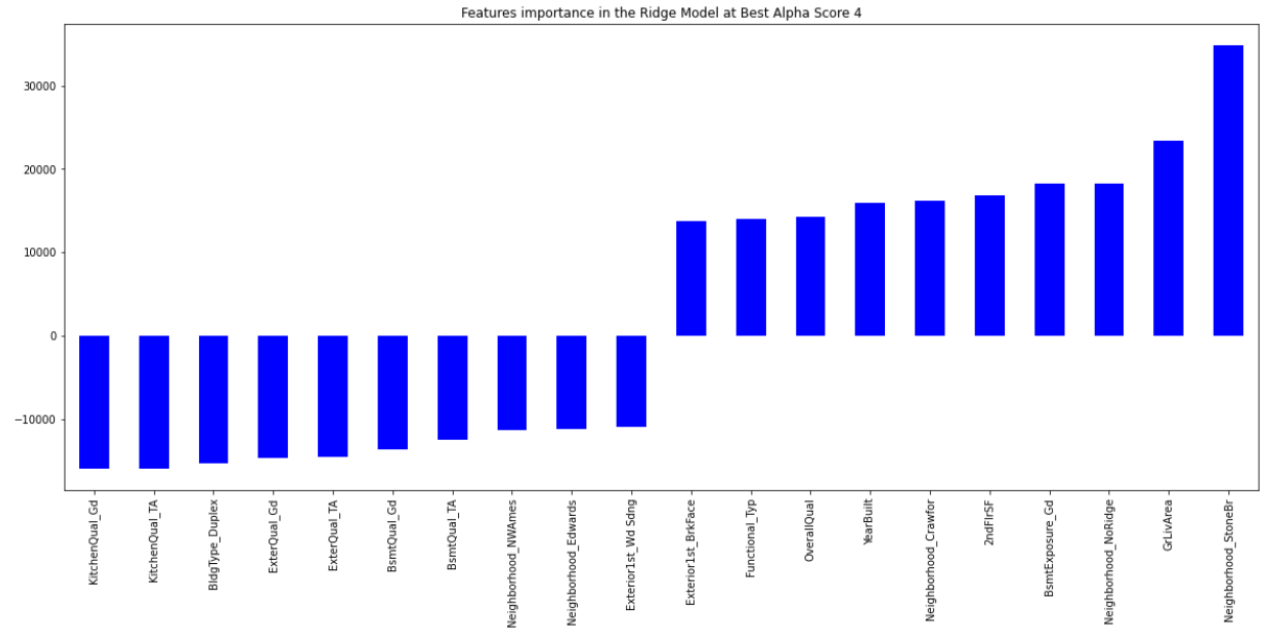
**Question 1:**
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

**Answer:**
The optimal value of alpha for ridge regression was **4** and for lasso regression was **100 .**

| Ridge Model Alpha=4 | Ridge Model Alpha=8 | Lasso Model Alpha=100 | Lasso Model Alpha=200 |
|---|---|---|---|
| 2ndFlrSF | Neighborhood_NoRidge | Neighborhood_Crawfor | **YearBuilt** |
| BsmtExposure_Gd | **OverallQual** | SaleCondition_Partial | **BsmtExposure_Gd** |
| Neighborhood_NoRidge | BsmtExposure_Gd | Neighborhood_NoRidge | SaleCondition_Partial |
| GrLivArea | GrLivArea | GrLivArea | GrLivArea |
| Neighborhood_StoneBr | Neighborhood_StoneBr | Neighborhood_StoneBr | Neighborhood_StoneBr |

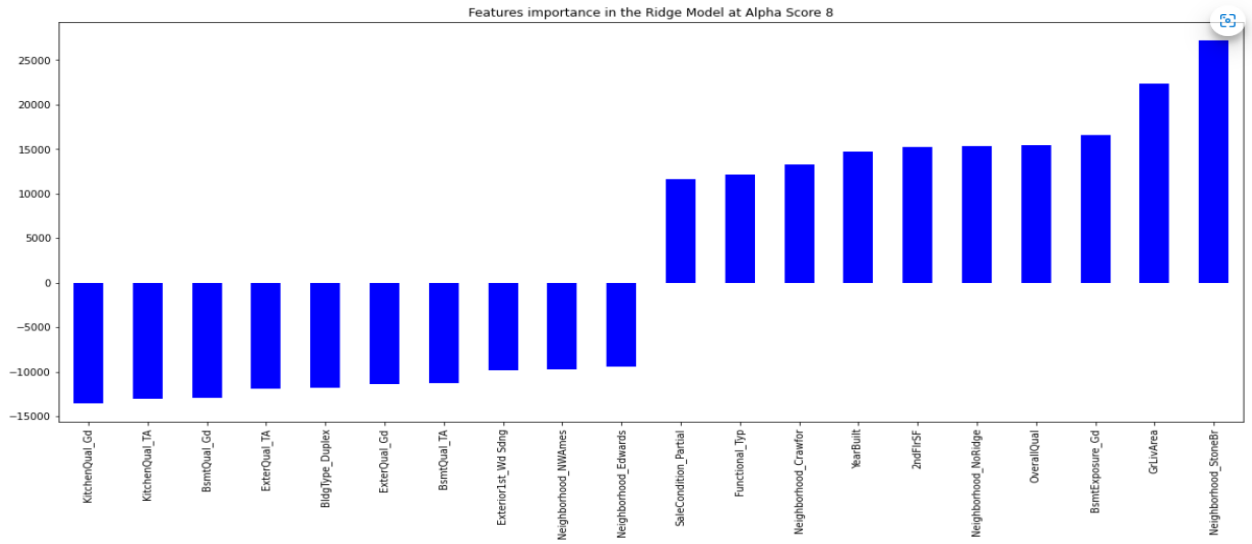
In above table if we change Ridge alpha 4 to 8 then in top 5 predictor variable **Overallqual** is important whereas in Lasso if we change alpha to 200 then important predictor variable are **YearBuilt,BsmtExposure_Gd.**

1. **Ridge Regression:**
   a. **Alpha = 4:-**




Ridge Regression - Chart – 1.1.1

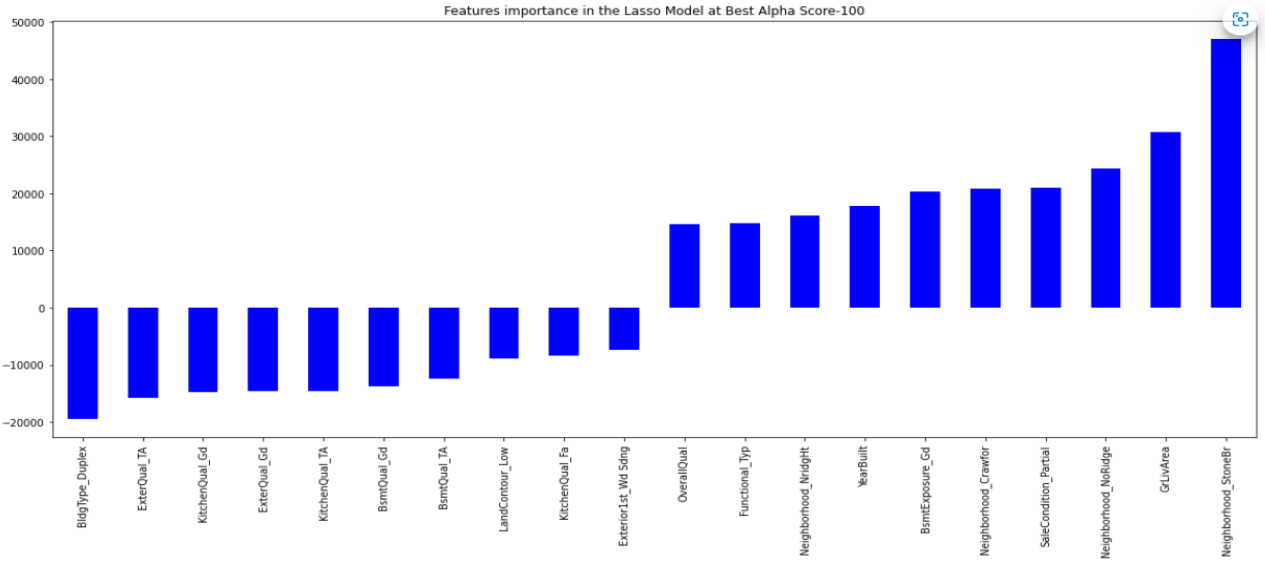
**b. Alpha = 8:-**




Ridge Regression - Chart – 1.1.2

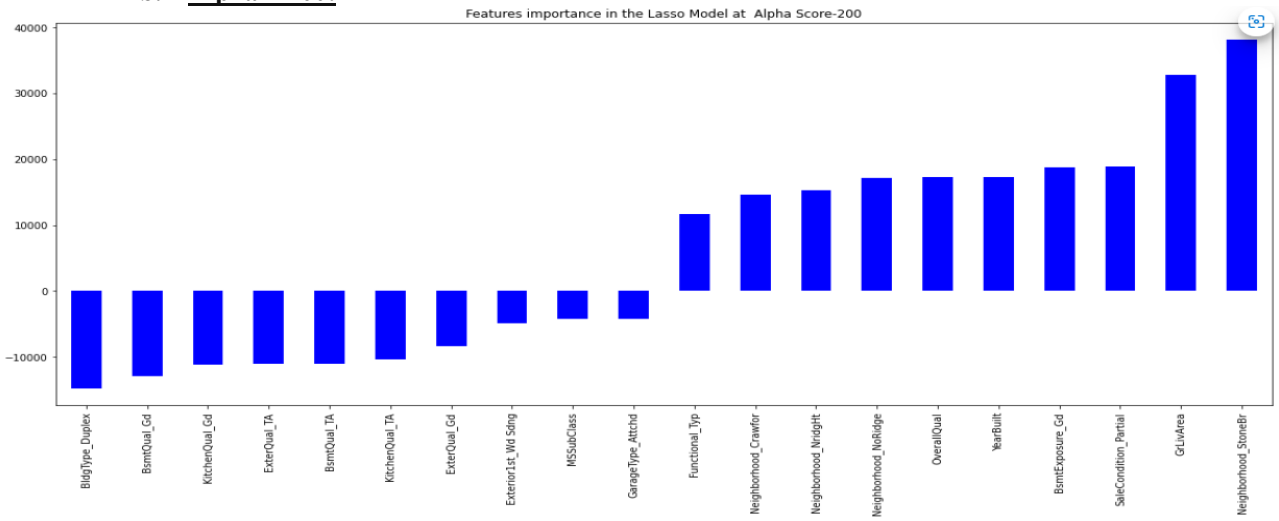
## 2. Lasso Regression:

a. **Alpha = 100:-**




Lasso Regression - Chart – 1.2.1

**b. Alpha = 200:**

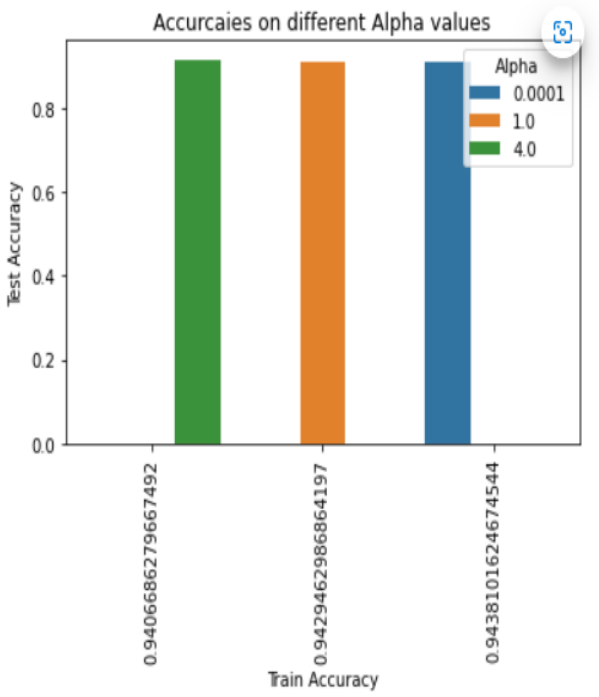



Lasso Regression - Chart – 1.2.2

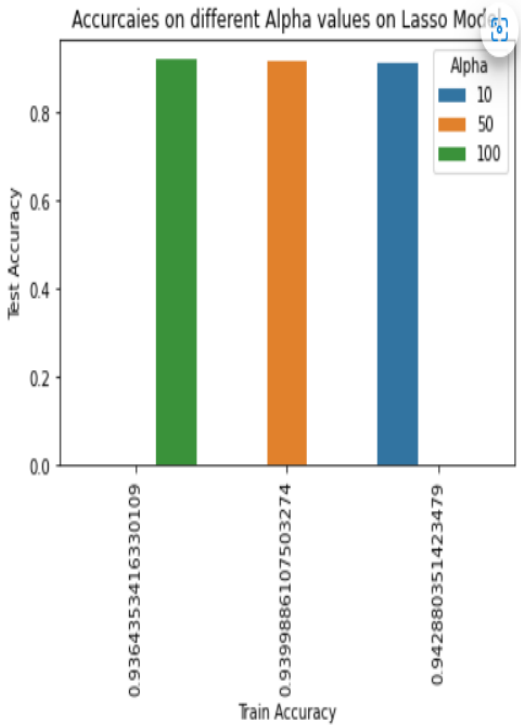
**Question 2:**
You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

**Answer:**
We will choose Lasso, because the r2_score is more generalised in Lasso as compared to Ridge Regression

**Ridge Regression:**




**Lasso 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 predictor variables are:

1. Functional_Typ
2. Neighborhood_NridgHt
3. YearBuilt
4. BsmtExposure_Gd
5. OverallQual

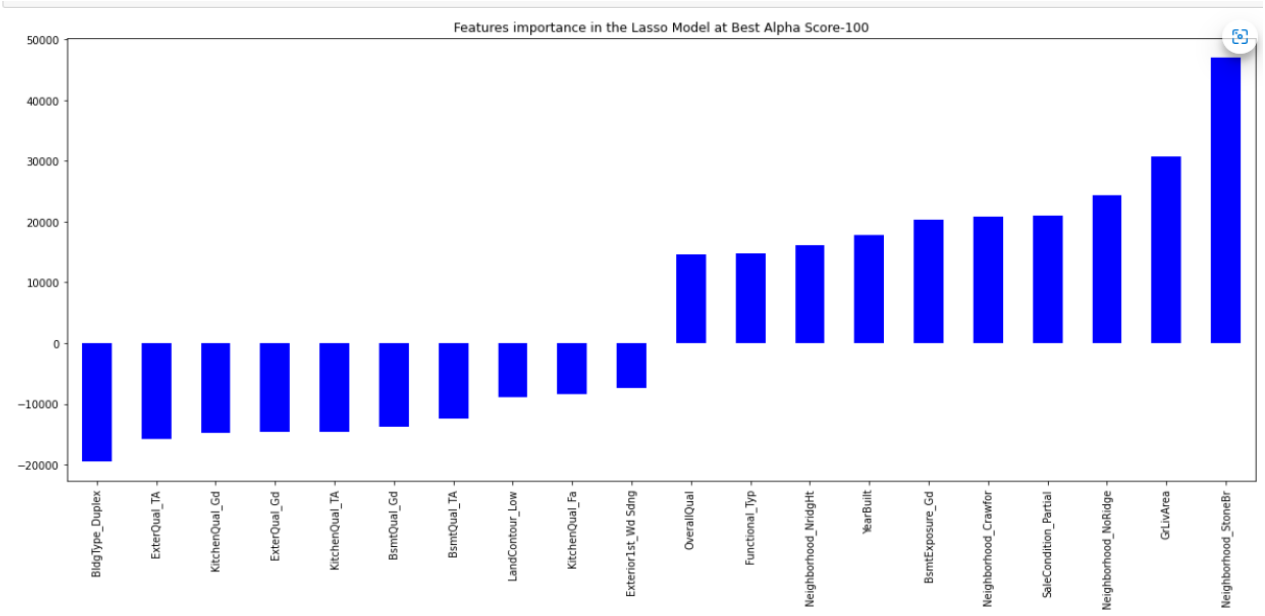



Chart – 3.1

**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 have removed the outliers from this data and the model's test accuracy is nearer to train accuracy hence we can state that the model is more generalise for doing the prediction.

- **Ridge Regression:**

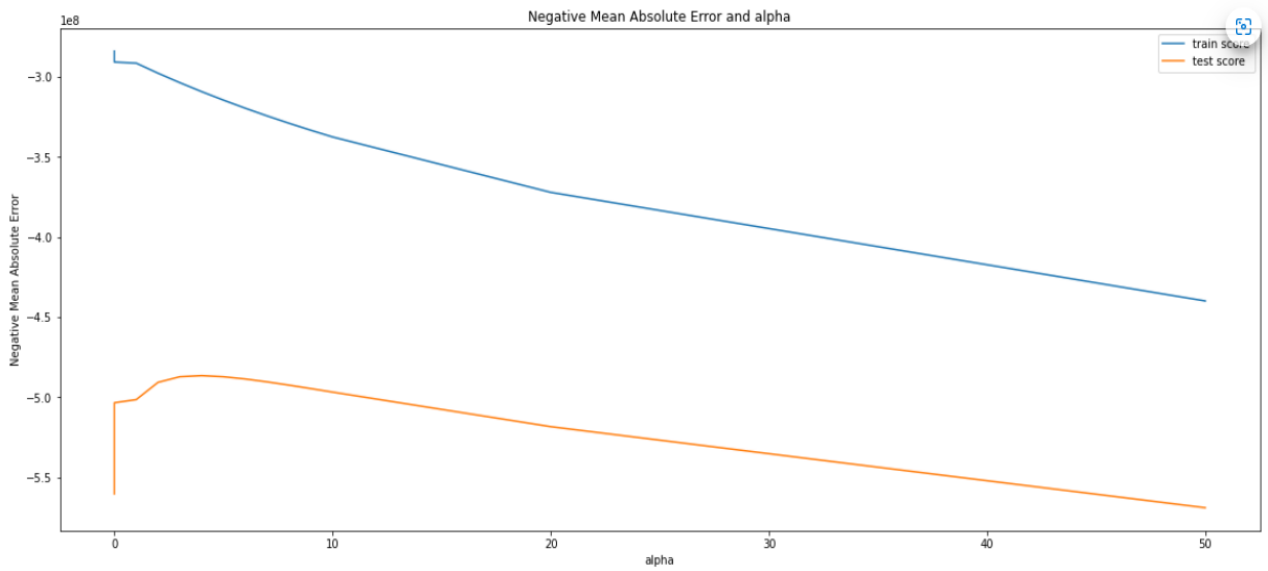



Chart – 4.1

- **Lasso Regression**

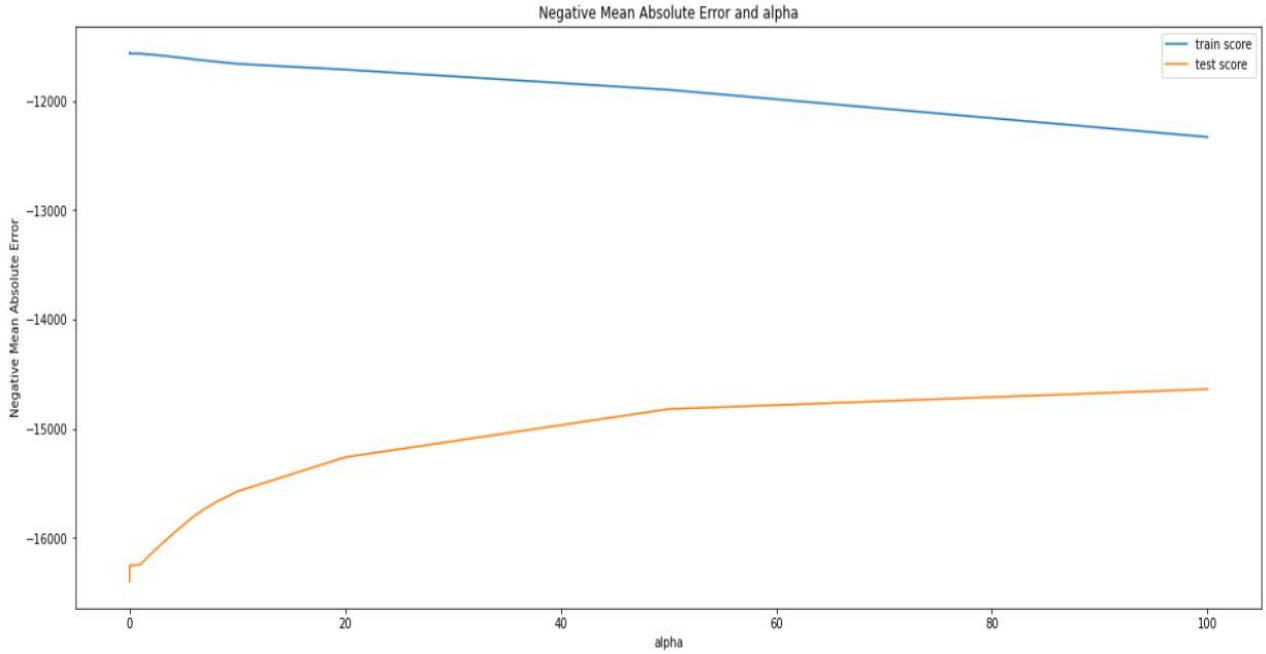



Chart – 4.2

For chart 4.1, we have seen that train and test accuracy are decreasing as we are increasing alpha. Whereas for chart 4.2, changes are not happening, which means lasso is more robust and generalize model.