**Antony John Sundar - https://github.com/AntonyJohnSundar/HousePrice-RidgeAndLasso.git**

**Advance Regression – Ridge and Lasso Regression Model Subjective Q & A**

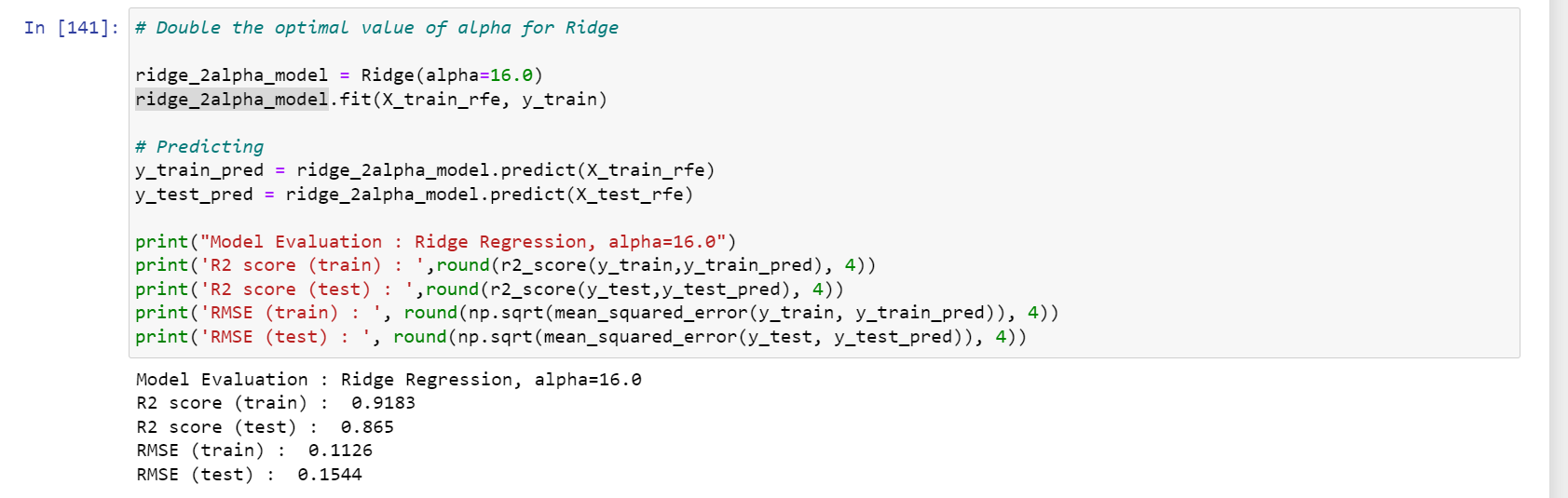
**Question 1**

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?

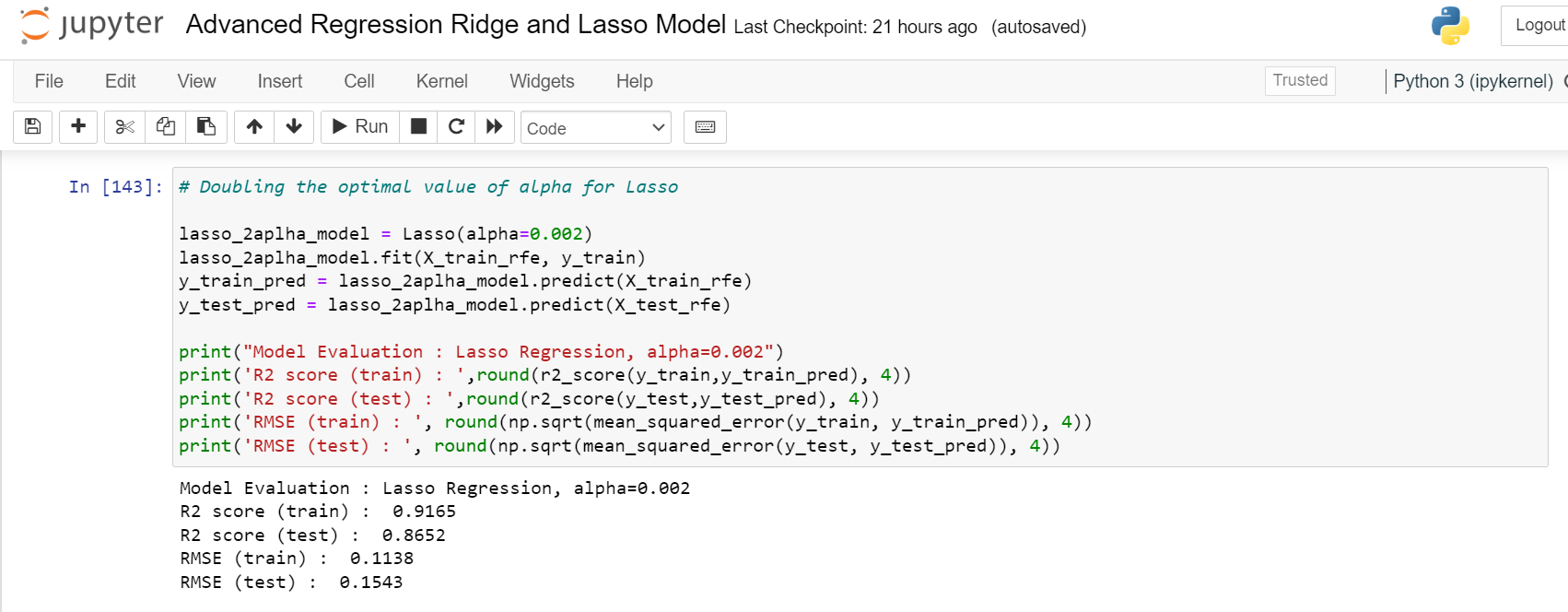
**Answer**:

1. **Optimal value of alpha for ridge and lasso regression are 8.0 and 0.001 respectively**
2. **Doubling the alpha values of ridge and lasso**

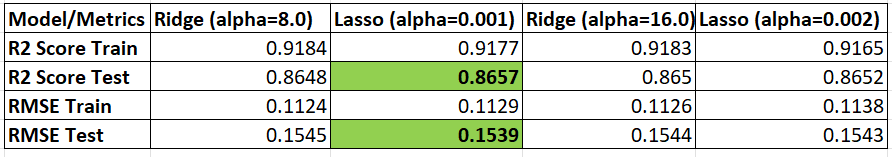
Metrics of ridge model when **alpha=16.0** (Optimal alpha =8.0)



Metrics of lasso model when **alpha=0.002** (Optimal alpha =0.001)



The metrics with optimal and doubled alpha values for Ridge and Lasso Model



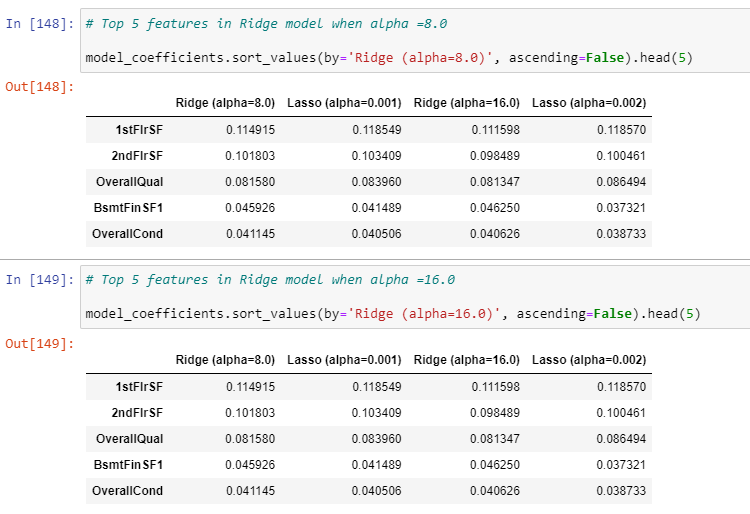
After doubling the alpha values of changes in the evaluation metrics are very minimal, with following observation

1. The Lasso model performed slightly lower on training data when alpha doubled
2. The Regression model does not show much of performance variation when alpha doubled
3. **The Most Important Predictors after doubling the alpha for ridge and lasso**

The most important predictors with optimal alpha value for ridge and lasso are similar to the most important predictors with the optimal alpha value doubled for ridge and lasso model

**Important Predictors - 1stFlrSF, 2ndFlrSF, OverallQual, BsmtFinSF1, OverallCond**

However, when the alpha value is doubled for lasso model, the coefficient of **OverallCond outperformed BasmtFinSF**1

****

****

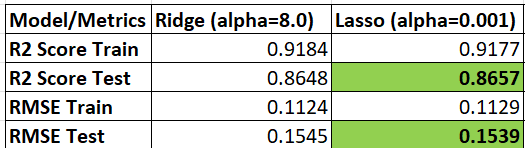
**Question 2**

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**

Optimal value of alpha for ridge and lasso regression are 8.0 and 0.001 respectively

The metrics with optimal values for Ridge and Lasso Model



Though the differences in performance metrics are very minimal, Lasso gets little edge over Ridge as it has slightly **better R2 Score and lesser RMSE on test data.**

Hence, our **choice of model** will be **Lasso Regression 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 are

1. 1stFlrSF
2. 2ndFlrSF
3. OverallQual
4. BsmtFinSF1
5. OverallCond

Model building dropping the above most important features



**The most important variables excluding the five most important predictor variable are**

1. **FullBath**
2. **KitchenQual**
3. **FirePlaces**
4. **GarageArea**
5. **LotArea**

**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 made sure the model is robust and generalized by using regularization. We have tried two of the regularization techniques, Ridge and Lasso in our model to see which suits better. We have seen that the Lasso model outperformed the Ridge in our model. The implications of the regularizations are

1. Mitigating overfitting - Regularization ensures that our model doesn’t memorize noise in the training data and gives a balance between bias and variance.
2. Feature selection – Regularization helped in identifying the most relevant features for prediction, thus we have simpler model with fewer parameters.

Moreover, we have used cross validation to ensure the model is performing well with test data

**In summary, regularization used in our model ensures robustness and generalizability by preventing overfitting, improving model performance, and aiding feature selection.**