# 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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

## Answer:

- Optimal value of lambda for **Ridge** regression is **100**
- Optimal value of lambda for **Lasso** regression is **0.01**

Doubling the values of lambda for **Ridge** and **Lasso**, we get **200** and **0.02** lambdas respectively. When we double the lambda values, we get the following R2 – Scores.

**RIDGE:**

R2-Score (Train set) for lambda 100 is = 0.9331470661244234

R2-Score (Test set) for lambda 100 is = 0.9006170818621794

R2-Score (Train set) for lambda 200 is = 0.927134598231389

R2-Score (Test set) for lambda 200 is = 0.8983194400823644

**LASSO:**

R2-Score (Train set) for lambda 0.01 is = 0.9222593845661436

R2-Score (Test set) for lambda 0.01 is = 0.9004034223356157

R2-Score (Train set) for lambda 0.02 is = 0.9046150295145502

R2-Score (Test set) for lambda 0.02 is = 0.8841894811355179

It is evident that as the lambda values get doubled for both Ridge and Lasso, the R2-Scores dropped slightly. Also, their variance shrinks as regularisation is doubled. R2 Scores dropped due to the compromised bias for the shrinking of variance.

**The variables that are significant for lambda 200 in Ridge regression are:**

- GrLivArea (0.1767405)
- OverallQual (0.1273205)
- TotalBsmtSF (0.1181521)
- BsmtFinSF1 (0.08905108)

- Neighborhood_NridgHt (0.07887876)
- GarageCars (0.07740398)
- BsmtExposure_Gd (0.06977058)
- LotArea (0.06700443)
- BsmtQual_Gd (-0.0704409)
- KitchenQual_Gd (-0.06082158)

**The variables that are significant for lambda 0.02 in Lasso regression are:**

- GrLivArea (0.344004)
- OverallQual (0.213223)
- BsmtFinSF1 (0.113270)
- SaleType_New (0.097687)
- Neighborhood_NridgHt (0.085888)
- TotalBsmtSF (0.080402)
- GarageCars (0.068784)
- BsmtExposure_Gd (0.065936)
- Neighborhood_StoneBr (0.053108)
- LotArea (0.050438)
- Age_from_YearBuilt (-0.055718)
- MSSubClass (-0.056723)

## 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:

**RIDGE:**

R2-Score (Train set) for lambda 100 is = 0.9331470661244234

R2-Score (Test set) for lambda 100 is = 0.9006170818621794

**LASSO:**

R2-Score (Train set) for lambda 0.01 is = 0.9222593845661436

R2-Score (Test set) for lambda 0.01 is = 0.9004034223356157

The R2 Score of **Ridge regression** is observed to be **slightly greater** than that of Lasso regression. But Ridge regression shrinks the **coefficients towards 0 but not exactly 0,** which implies that all the 231 features are included in the Ridge model. When the dataset has a greater number of features, it is tedious and time consuming to use Ridge regression as it includes all the features in the model. As the number of features increase, the **model becomes too complex**. As the complexity increases, the **variance also increases but the bias compromise is reduced.**

On the other hand, **Lasso regression** pushes **some of the coefficients to exactly 0**, which implies it does **feature selection**. Here, with the optimal value of lambda **0.01**, Lasso regression selected **88 significant features** while the Ridge regression contains all the (231) features. As the number of features are mitigated, the model is not as complex as Ridge and the variance is also shrinked with reasonable bias compromise.

Hence, it is better to use **Lasso regression** as it does **feature selection** and thereby **reducing the complexity** of the 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 5 top-most variables that are significant for lambda 0.01 in Lasso regression are:

- GrLivArea (0.356475)
- OverallQual (0.174764)
- SaleType_New (0.104234)
- BsmtFinSF1 (0.098615)
- Age_from_YearBuilt (-0.092890)

Suppose these predictor variables are not present in the incoming data. Now, we will create another model in Lasso regression with lambda 0.01 and predict the new significant features.

The significant features of the new model are:

- TotalBsmtSF (0.380835)
- 2ndFlrSF (0.241000)
- GarageCars (0.106787)
- BsmtQual_Gd (-0.147367)
- KitchenQual_TA (-0.144912)
- KitchenQual_Gd (-0.133956)
- BsmtQual_TA (-0.108041)
- SaleCondition_Partial (0.093914)
- BsmtExposure_Gd (0.080917)
- Age_from_YearBuilt (-0.087426)

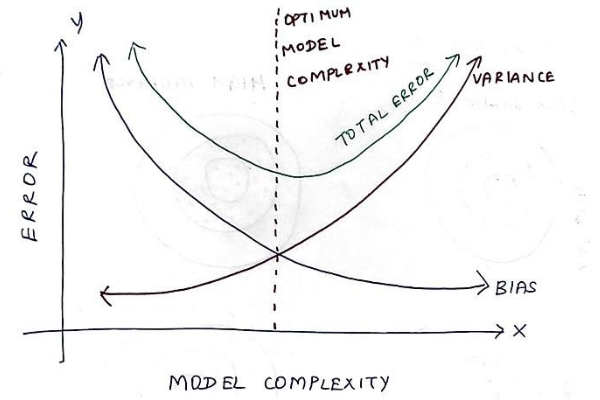
## Question 4

## 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:

When the model is simple, the variance is low, but the bias is high. If the model works well on the train set as well as the test set, it is said to be robust (i.e., it is neither underfitted nor overfitted). Simple models can be easily generalisable though the bias is reasonably compromised. A model should be as simple as necessary but not simpler than that.

There is always a trade-off between bias and variance.




BIAS VARIANCE TRADE-OFF

**BIAS**

→ Quantifies how much accurate the model is likely to be on future data.

→ High for simple models, Low for complex models.

**VARIANCE**

→ How sensitive is the model to input data?

→ Variance is the degree of changes reflected on some test data with respect to the changes in the training data.

→ Low for simple models, High for complex models

Expected error that a model makes = BIAS + VARIANCE

**Optimum model complexity:** Roughly where bias and variance meet each other.

Extremely simple models (UNDERFITTED) are likely to fail in predicting complex real-world phenomena. Extremely complex models (OVERFITTED) do not work well on test data or unseen data.

To get the correct balance (robustness) between bias and variance or simplicity and complexity, we use **regularisation techniques** like Ridge or Lasso, which prevents the model being complex unnecessarily.