# Assignment Part-II

**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?

The optimal value for **Ridge is 0.8** and **Lasso is 0.0003**

| | Features | rfe_support | rfe_ranking | Coefficient |
|---|---|---|---|---|
| 3 | GrLivArea | True | 1 | 0.7546 |
| 0 | OverallQual | True | 1 | 0.5373 |
| 1 | OverallCond | True | 1 | 0.3889 |
| 7 | MSZoning_FV | True | 1 | 0.2468 |
| 16 | BsmtFullBath_3 | True | 1 | 0.2216 |
| 2 | BsmtFinSF1 | True | 1 | 0.2170 |
| 8 | MSZoning_RH | True | 1 | 0.1331 |
| 18 | GarageCars_4 | True | 1 | 0.1288 |
| 9 | Neighborhood_Crawfor | True | 1 | 0.1264 |
| 10 | RoofStyle_Mansard | True | 1 | 0.1031 |

| | Features | rfe_support | rfe_ranking | Coefficient |
|---|---|---|---|---|
| 5 | GrLivArea | True | 1 | 0.685697 |
| 1 | OverallQual | True | 1 | 0.418675 |
| 4 | TotalBsmtSF | True | 1 | 0.341698 |
| 2 | OverallCond | True | 1 | 0.315619 |
| 3 | BsmtFinSF1 | True | 1 | 0.157505 |
| 28 | Exterior1st_BrkFace | True | 1 | 0.102977 |
| 0 | LotArea | True | 1 | 0.102770 |
| 18 | Neighborhood_Crawfor | True | 1 | 0.099206 |
| 20 | Neighborhood_StoneBr | True | 1 | 0.096606 |
| 41 | GarageType_Attchd | True | 1 | 0.077496 |

The optimal value for **Ridge is 1.6** and **Lasso is 0.0006**

| | Features | rfe_support | rfe_ranking | Coefficient |
|---|---|---|---|---|
| 3 | GrLivArea | True | 1 | 0.6830 |
| 0 | OverallQual | True | 1 | 0.5316 |
| 1 | OverallCond | True | 1 | 0.3725 |
| 2 | BsmtFinSF1 | True | 1 | 0.2170 |
| 7 | MSZoning_FV | True | 1 | 0.2092 |
| 16 | BsmtFullBath_3 | True | 1 | 0.1550 |
| 9 | Neighborhood_Crawfor | True | 1 | 0.1265 |
| 18 | GarageCars_4 | True | 1 | 0.1124 |
| 11 | Exterior1st_BrkFace | True | 1 | 0.1018 |
| 10 | RoofStyle_Mansard | True | 1 | 0.0948 |

| | Features | rfe_support | rfe_ranking | Coefficient |
|---|---|---|---|---|
| 5 | GrLivArea | True | 1 | 0.683673 |
| 1 | OverallQual | True | 1 | 0.470830 |
| 4 | TotalBsmtSF | True | 1 | 0.357916 |
| 2 | OverallCond | True | 1 | 0.302526 |
| 3 | BsmtFinSF1 | True | 1 | 0.161862 |
| 18 | Neighborhood_Crawfor | True | 1 | 0.096135 |
| 0 | LotArea | True | 1 | 0.094840 |
| 28 | Exterior1st_BrkFace | True | 1 | 0.091449 |
| 20 | Neighborhood_StoneBr | True | 1 | 0.075460 |
| 16 | MSZoning_RL | True | 1 | 0.054638 |

After doubling the values, the coefficient values are decreased slightly compared to previous one. Some of the Top features are changed in Ridge Regression and Lasso Regression.

**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?

I will choose Lasso regression instead of Ridge. In Lasso, the feature selection is automatic. IF number of variables increases, we prefer Lasso Regression.

**R2 value:**

In Ridge – Train = 0.918885 After Double Ridge Train – 0.916779

Test = 0.902600 Test - 0.90214

In Lasso – Train = 0.934936 After Double, Lasso Train – 0.928608

Test = 0.906698 Test - 0.910349

| | Metric | Linear Regression | Ridge Regression | Lasso Regression |
|---|---|---|---|---|
| 0 | R2 Score (Train) | 0.920434 | 0.918885 | 0.934936 |
| 1 | R2 Score (Test) | 0.899211 | 0.902600 | 0.906698 |
| 2 | RSS (Train) | 12.349095 | 12.589498 | 10.098233 |
| 3 | RSS (Test) | 6.060608 | 5.856842 | 5.610428 |
| 4 | MSE (Train) | 0.111349 | 0.112428 | 0.100692 |
| 5 | MSE (Test) | 0.118997 | 0.116980 | 0.114492 |

Also, R squared value, RSS and MSE for Lasso regression is higher compared to Ridge 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?

Before removing the top 5 Predictor variables in the lasso model.

| | Features | rfe_support | rfe_ranking | Coefficient |
|---|---|---|---|---|
| **5** | GrLivArea | True | 1 | 0.683673 |
| **1** | OverallQual | True | 1 | 0.470830 |
| **4** | TotalBsmtSF | True | 1 | 0.357916 |
| **2** | OverallCond | True | 1 | 0.302526 |
| **3** | BsmtFinSF1 | True | 1 | 0.161862 |
| **18** | Neighborhood_Crawfor | True | 1 | 0.096135 |
| **0** | LotArea | True | 1 | 0.094840 |
| **28** | Exterior1st_BrkFace | True | 1 | 0.091449 |
| **20** | Neighborhood_StoneBr | True | 1 | 0.075460 |
| **16** | MSZoning_RL | True | 1 | 0.054638 |

After removing the top 5 predictor variables in the Lasso regression. Below variables are most important.

| | Features | rfe_support | rfe_ranking | Coefficient |
|---|---|---|---|---|
| **3** | 2ndFlrSF | True | 1 | 0.371099 |
| **4** | GarageArea | True | 1 | 0.169599 |
| **40** | GarageType_BuiltIn | True | 1 | 0.098220 |
| **31** | BsmtExposure_Gd | True | 1 | 0.072957 |
| **0** | LotArea | True | 1 | 0.071065 |
| **19** | Neighborhood_Somerst | True | 1 | 0.054093 |
| **12** | MSZoning_RL | True | 1 | 0.044602 |
| **1** | MasVnrArea | True | 1 | 0.043464 |
| **30** | Exterior2nd_Wd Sdng | True | 1 | 0.039491 |
| **45** | GarageCars_3 | True | 1 | 0.037321 |

**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?

Accuracy:

Model will not impact by outliers in the training data and test accuracy is not lesser than training.

It is used to identify the relationships and patterns between the variables in the data.

The difference between the Train and Test R squared values for the given model is lies between 5% accuracy range. We can say that the model is robust and generalisable.