Q.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:

- Optimal value of lambda for Ridge Regression = **10**
- Optimal value of lambda for Lasso = **0.001**

The optimal value of alpha for Ridge and Lasso regression depends on the specific dataset and problem you are working with. Generally, alpha is a hyperparameter that controls the amount of regularization in the models. Higher values of alpha lead to stronger regularization, which can help prevent overfitting but may also result in underfitting if the value is too high.

To find the optimal value of alpha, you can use techniques like cross-validation to evaluate different alpha values and choose the one that results in the best performance on your validation data.

If you choose double the value of alpha for both Ridge and Lasso regression, the regularization strength will increase. For Ridge regression, this means the impact of the L2 regularization term will be stronger, while for Lasso regression, the impact of the L1 regularization term will be stronger.

In Ridge regression, larger alpha values lead to coefficients being closer to zero but not exactly zero. So, increasing alpha in Ridge regression will continue to shrink the coefficients towards zero, but none of them will be eliminated completely.

In Lasso regression, however, larger alpha values can lead to some coefficients becoming exactly zero. This means that some features become completely irrelevant to the model, and they will be effectively removed from the model. Lasso is particularly useful for feature selection because it can automatically select the most important predictors and discard the less important ones.

Q.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

- The model we will choose to apply will depend on the use case.
- If we have too many variables and one of our primary goal is feature selection, then we will use **Lasso**.
- If we don't want to get too large coefficients and reduction of coefficient magnitude is one of our prime goals, then we will use **Ridge Regression**.

Q.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

After dropping our top 5 lasso predictors, we get the following new top 5 predictors:-

- `2ndFlrSF`
- `Functional_Typ`
- `1stFlrSF`
- `MSSubClass_70`
- `Neighborhood_Somerst`

Q.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

- A model is **robust** when any variation in the data does not affect its performance much.
- A **generalizable** model is able to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.
- To make sure a model is robust and generalizable, we have to **take care it doesn't overfit**. This is because an overfitting model has very high variance and a smallest change in data affects the model prediction heavily. Such a model will identify all the patterns of a training data, but fail to pick up the patterns in unseen test data.
- In other words, the model should not be too complex in order to be robust and generalizable.

- If we look at it from the prespective of **Accuracy**, a too complex model will have a very high accuracy. So, to make our model more robust and generalizable, we will have to decrease variance which will lead to some bias. Addition of bias means that accuracy will decrease.

- In general, we have to find strike some balance between model accuracy and complexity. This can be achieved by Regularization techniques like Ridge Regression and Lasso.