

## Subjective Questions

**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?

**Ans.** The optimum value of **alpha for Ridge is 20** and for **Lasso is 0.001**. So in my model the changes that followed after choosing double the value of alpha for both ridge and lasso are:

- The train and test set accuracy score for Lasso went down by 1%.
- More importantly, both the model's coefficient value of some/all predictor variables have changed.

The most important predictor variables for Ridge and Lasso after implementing these changes are:

Model	Important Predictor Variables
Lasso	1. TotAreaSF (Derived variable) 2. OverallQual
Ridge	1. TotAreaSF (Derived variable) 2. GrLivArea

---

**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?

**Ans.** I would choose Lasso regression because Lasso helps in feature elimination and helps with better model readability and interpretability.

---

**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?

**Ans.** After removing first 5 variables, we have the following important variables:

- Neighborhood\_StoneBr**
- MSZoning\_RL**
- MSZoning\_RM**
- MSZoning\_RH**
- FullBath**

---

**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?

**Ans.** One of the ways by which a model can be made more robust and generalisable is by removing the outliers which are extreme values and does not hold any relevance to the data. The model should be robust and should not be influenced by outliers in the data. We should make sure that our model is generalisable so that it does not overfits which means that it should not show high accuracy in training data while poor accuracy score in test data. Also, it is necessary to for our model to be trusted for predictive analysis.

In order to achieve that we need to use robust methods in our models. For e.g., Robust Scaler, transforming the independent variables and/or dependent variable are some of the ways to get a robust and reliable model. In this way we get good model metrics in the train data as well as in the test data. The reason behind this is when we get a robust and generalisable model, we make sure that the model doesn't change a lot when the data or some of the parameters like some variables are changing. Hence this contributes to better model accuracy.