Subjective Questions Solution

Q1) 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?

Ans)

- The optimal value of alpha for *Ridge regression is 10.0* and for *Lasso regression is 0.0002*.
- If we double the value of alpha for *Ridge regression* then the *train accuracy* of model has *decreased* from *90% to 88%* and the *test accuracy* of the model has *decreased* from *85% to 84%.*
- Similarly if we double the value of alpha for *Lasso regression* then the *train accuracy* of the model has *decreased* from *90.1% to 86.9%* and the *test accuracy* of the model has *decreased* from *85.7% to 85.2%.*
- After the change is implemented the most important predictor variables from the Ridge regression are *'OverallQual_9', 'TotRmsAbvGrd', 'GrLivArea', 'GarageCars', 'OverallQual_10'*.
- After the change is implemented the most important predictor variables from the Lasso regression are *'GrLivArea', 'OverallQual_9', 'OverallQual_10', 'GarageCars', 'Neighborhood_NoRidge'*.

Q2) 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) We will choose the Lasso regression because of following reasons:

- Lasso model has better accuracy(Train: 90.1%, Test: 85.7%) than Ridge model(Train: 90%, Test: 85%).
- Lasso model also eliminates the less important features from the model, and hence the model becomes simpler.

Q3) After building the model, you realized 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 the most important predictor variables from the lasso model the new model has following most important predictors variables:

- 1stFlrSF
- 2ndFlrSF
- GarageArea
- Neighborhood_NridgHt
- BsmtFullBath

Q4) How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Ans) Ensuring model robustness and generalizability involves following several steps:

- **Data Quality**: Start with clean, representative data to train the model.
- **Feature Selection**: Choose relevant features that capture the underlying patterns in the data.
- **Regularization**: Employ techniques like Ridge/Lasso regularization to prevent overfitting.
- **Cross-Validation**: Validate the model on multiple subsets of the data to ensure consistency.
- **Ensemble Methods**: Combine predictions from multiple models to improve generalizability.
- **Testing on diverse datasets**: Evaluate the model's performance on different datasets to ensure it can handle various scenarios.

Implications for accuracy: Prioritizing robustness and generalizability may slightly decrease the model's accuracy on the training data but often leads to better performance on unseen data. It's a trade-off between fitting the training data perfectly and being able to make accurate predictions on new, unseen data.
