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 alpha values you provided for Ridge and Lasso are:

- Ridge: $\alpha=0.5689866029018293$
- Lasso: $\alpha=0.00014563484775012445$

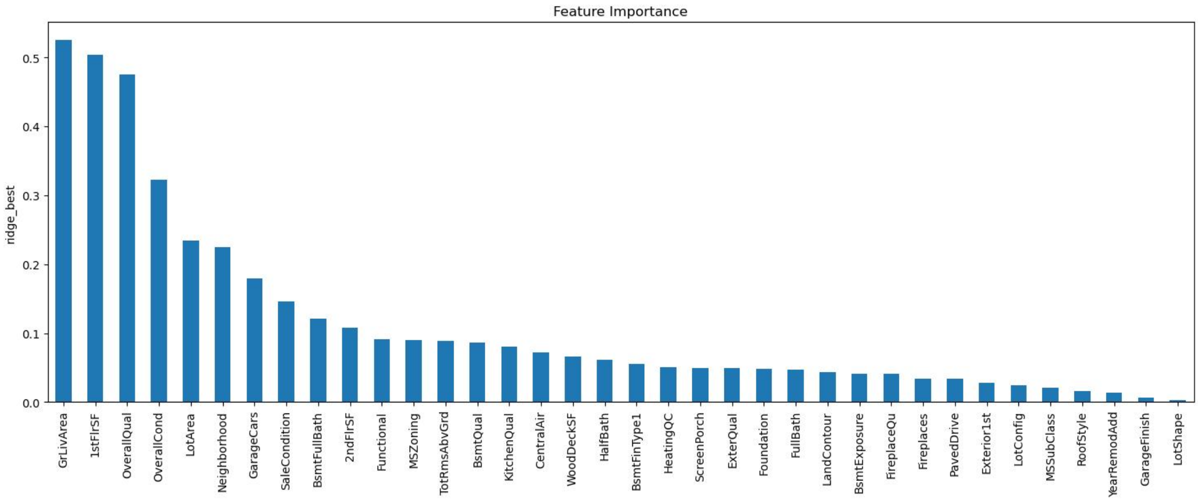
Ridge Regression

- **Increased Regularization:** Doubling alpha in Ridge regression increases the penalty on large coefficients.
- **Effect on R2:**
  - The R2 value may decrease because the model will underfit slightly as regularization increases.
  - Ridge doesn't zero out coefficients but shrinks them, so all predictors still contribute.
- **Change in Coefficients:**
  - Smaller coefficients for less significant predictors.
  - Important variables remain influential but with smaller magnitudes.

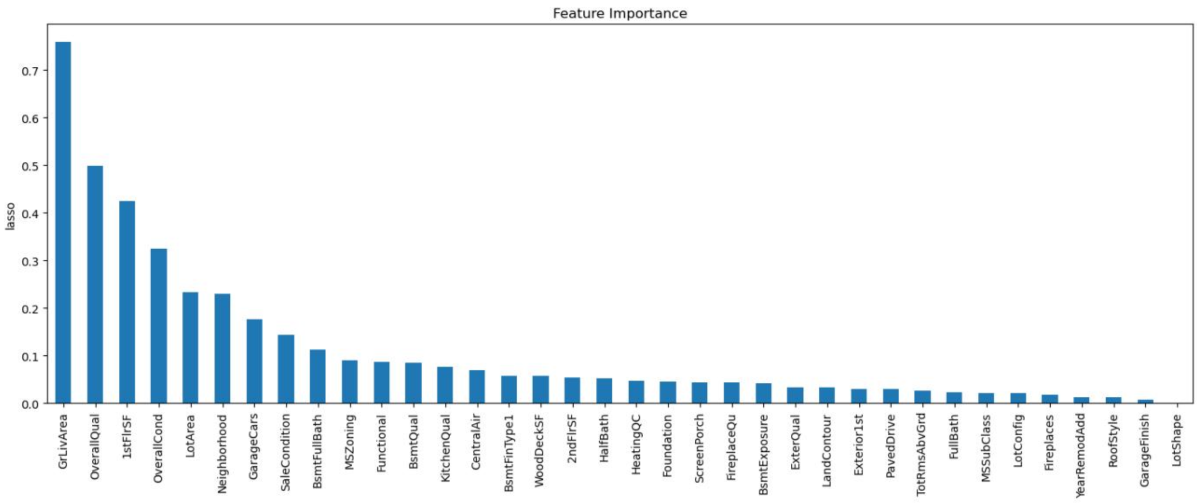
Lasso Regression

- **Increased Sparsity:** Doubling alpha in Lasso increases the L1L_1L1 -norm penalty, which leads to more coefficients being driven to zero.
- **Effect on R2:**
  - R2 will reduce more significantly compared to Ridge because Lasso might zero out some predictors, reducing model complexity.
  - If alpha becomes too large, Lasso may underfit by retaining only a few predictors.
- **Change in Coefficients:**
  - Fewer variables will have nonzero coefficients.
  - Important predictors with the strongest relationships to the target will remain, while weaker predictors will be removed.

Ridge Regression



Lasso Regression



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?

Choosing Ridge Regression is a sound decision when the train and test R2 values are close, as this indicates the model generalizes well without overfitting or underfitting. Here's an explanation of why Ridge is a better choice for your case and the rationale behind it:

Consistency of Test and Train R2R^2R2:

- Ridge regression works well when all predictors contribute to the outcome and multicollinearity is present. It shrinks the coefficients but retains all predictors, providing stability in the model.
- The close R2R^2R2 values between train and test data suggest that Ridge has effectively handled variance without underfitting.

Behavior of Lasso:

- Lasso introduces sparsity by zeroing out less important coefficients. This can be an advantage in high-dimensional datasets with many irrelevant predictors.
- However, if all predictors are valuable or the relationships are complex, Lasso's aggressive regularization might lead to underfitting or loss of critical information.

Stability with Multicollinearity:

- Ridge handles multicollinearity better than Lasso by distributing the penalty evenly across correlated predictors, rather than selecting one and shrinking the rest to zero.

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?

Variables and coefficient

- GrLivArea - 0.758292
- OverallQual - 0.498501
- 1stFlrSF - 0.424401
- OverallCond - 0.324838
- LotArea - 0.232544

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?