

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?

Ans: predictor variables after we double the alpha values are: Neighborhood_ClearCr inf Exterior2nd_BrkFace inf Neighborhood_Veenker inf Neighborhood_Somerst inf BsmtExposure inf

Question2: 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: Model to be chosen depend on a use case. Lasso is to be used if there are too many variables and our goal is feature selection Ridge Regression is to be used if no large coefficients and reduction of coefficient value is goals.

Question3: 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: `betas['Lasso'].sort_values(ascending=False)[:5]`

```
Neighborhood_NoRidge      58906.988883
Neighborhood_NridgHt      51633.819474
Neighborhood_Somerst      32800.914057
OverallQual                27267.629475
1stFlrSF                  26538.579429
Name: Lasso, dtype: float64
```

Out [899] :

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?

Ans: Non-robust models can lead to severe consequences, such as financial losses or compromised safety. Achieving robustness often involves a trade-off with accuracy. While aiming for high accuracy on a specific dataset is tempting, it can lead to overfitting or lack of generalization to new data. It ensures that the model can handle diverse inputs or circumstances. Train and validate the model on the first two subsets and evaluate its performance on the unseen test data.

Robustness ensures reliability, while accuracy ensures good performance on the validation set.