

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?

There is no definitive answer to this question, as the optimal value of alpha will vary depending on the specific dataset and the desired goal of the regression. However, in general, a good starting point for alpha is 1.0. For ridge regression, increasing the value of alpha will typically result in a model with better predictive accuracy, but at the cost of increased complexity. For lasso regression, a higher alpha value will lead to a simpler model with fewer features, but may sacrifice some predictive accuracy. Ultimately, it is up to the user to experiment with different alpha values to find the optimal balance for their particular problem.

If you choose to double the value of alpha for both ridge and lasso, the changes in the model will be as follows:

- For ridge, the change will be minimal, as doubling alpha will simply add more weight to the already existing penalties for large coefficients.

- For lasso, however, doubling alpha will have a more significant effect. This is because lasso relies on setting coefficients to 0 in order to reduce overfitting, and doubling alpha will therefore increase the amount of coefficients that are set to 0. This will in turn lead to a model that is more simplified and easier to interpret.

Yearbuilt and overallqual are the most important variable after the change is implemented.

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?

We have tried with different lambda in the assignment for the lasso and ridge regression. We find the 0.1 as the optimal value that we used to train the model.

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?

The first and most important variables of the dataset that we used to build the lasso model are Yearbuilt , GrLivarea , Garage cars , lots shape and utilitues. We find that these variables will not be available in the incoming data. So the most important variable now is Overalqual , Garage area, 1stslrfs, Landcountur and full bath

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?

There are a few key things to keep in mind when trying to ensure that a regression model is robust and generalizable. First, it is important to use a large, representative dataset when training the model. This will help to ensure that the model is not overfitting to any particular quirks in the training data. Second, it is important to tune the model carefully, using cross-validation to ensure that it is not overfitting to the training data. Finally, it is important to evaluate the model on independent test data to get a realistic estimate of its performance.

The implications of the same for the accuracy of the model are not entirely clear. It is possible that the model is less accurate when applied to new data, but it is also possible that the model is more accurate when applied to new data. It is not clear why the model would be more or less accurate when applied to new data.