

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?

Answer 1 : The optimal value of alpha found through the model was , 50 for ridge and 0.001 for lasso. If we increase / double the alpha value, there will be more coefficients moving towards 0 for ridge and exactly 0 for lasso. Some variables like MSSubClass , BsmtUnfSF , EnclosedPorch , 3SsnPorch , MiscVal started to show as coefficient 0 when the alpha for lasso was doubled. After the changes the some important predictor variables that could be identified were GrLivArea , OverallQual , YearBuilt , LotArea , OverallCond , 1stFlrSF , GarageCars.

Question 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?

Answer 2 : Post doing a model evaluation , it was found that the lasso fit gave a slightly better r^2 score on the test data hence we can consider lasso regression as the primary choice. Lasso also reduces the number of predictors hence giving a simpler model without compromising on the model performance on test data.

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.95742679255587714859	0.95485863663479186858	0.95351563737130251575
1	R2 Score (Test)	-28771817118095445959114752.00000000000000000000	0.86162224253520003359	0.88137925992347476711
2	RSS (Train)	0.04020362281634595747	0.04262883759774083875	0.04389708678715933260
3	RSS (Test)	12250485372647249337647104.00000000000000000000	0.05891858295790786820	0.05050642561905232392
4	MSE (Train)	0.00627508660236144461	0.00646158221025495364	0.00655699690487482507
5	MSE (Test)	167239782008.97430419921875000000	0.01159815988376118948	0.01073831802546847645

Question 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?

Answer 3 : The 2nd set of top 5 predictor variables can be 1stFlrSF , GarageCars , YearRemodAdd , BsmtFinSF1 , KitchenQual .

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?

Answer 4 : To make sure the model is robust and generalizable, we need to ensure the model not only fits well on the training data, but it also must perform well on the test data. We need to do a trade off between bias and variance in a way that we can achieve optimal model complexity which would also mean having lowest error. To do this we add regularization techniques like Ridge and Lasso, where we add a penalty term to the cost function. On identifying the best hyperparameter, it can reduce the risk of overfitting and underfitting.