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 :

Alpha value for ridge is : 3

R2 = 0.8410106159376006

Alpha value for Lasso is : 0.0009

R2 = 0.8373792294285463

	Ridge Co-Efficient
TotRmsAbvGrd	87216.294857
1stFlrSF	80182.775599
Neighborhood_NoRidge	63886.575794
GarageArea	56840.806566
FullBath	44092.928115
RoofMatl_WdShngl	42917.122446
BsmtFullBath	42097.366887
LotArea	37430.763081
ScreenPorch	37148.900365
Exterior2nd_ImStucc	34078.131805
KitchenQual_Ex	33205.435459
BsmtQual_Ex	32883.590621
Neighborhood_StoneBr	31625.317633
Neighborhood_NridgHt	24855.092814

	Lasso Co-Efficient
RoofStyle_Shed	178640.149993
TotRmsAbvGrd	102454.833590
1stFlrSF	100211.541583
RoofMatl_WdShngl	89740.837418
RoofMatl_Metal	65025.943344
Exterior2nd_ImStucc	57453.356263
Neighborhood_NoRidge	56898.357766
GarageArea	55713.010927
LotArea	52872.631290
SaleType_Con	52681.804289
BsmtFullBath	47755.789101
RoofStyle_Mansard	45207.892691
ScreenPorch	41014.518805
FullBath	38970.549411
Condition2_PosA	37395.603618
OverallCond	35221.501831

After doubling Ridge and Lasso Alpha values :

Ridge R2 : R2 : 0.8366636670692804

Lasso R2 : 0.837380163237525

Very small changes in R2 changes. No change in predictor variable

	Ridge Doubled Alpha Co-Efficient
TotRmsAbvGrd	77032.696169
1stFlrSF	71420.442363
Neighborhood_NoRidge	62301.052977
GarageArea	54705.018464
FullBath	44867.961682
BsmtFullBath	38564.146979
BsmtQual_Ex	34126.839236
LotArea	34064.074697
KitchenQual_Ex	33963.092354
ScreenPorch	32811.285036
Neighborhood_StoneBr	30061.666319
RoofMatl_WdShngl	29398.101959
Exterior2nd_ImStucc	26659.937839
Neighborhood_NridgHt	24995.099148
WoodDeckSF	24787.311522
ExterQual_Ex	22770.901212
OpenPorchSF	22158.231472

	Lasso Doubled Alpha Co-Efficient
RoofStyle_Shed	178633.879627
TotRmsAbvGrd	102454.772465
1stFlrSF	100211.442027
RoofMatl_WdShngl	89740.472002
RoofMatl_Metal	65024.038930
Exterior2nd_ImStucc	57453.028915
Neighborhood_NoRidge	56898.601883
GarageArea	55713.119564
LotArea	52872.307397
SaleType_Con	52681.368854
BsmtFullBath	47755.691141
RoofStyle_Mansard	45205.620607
ScreenPorch	41014.465693
FullBath	38970.609256
Condition2_PosA	37394.016665
OverallCond	35221.330281
KitchenQual_Ex	31655.811351

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?

The optimal values for ridge and Lasso are :

- Ridge : 3
- Lasso : .0009
- The SME is almost same for both models.
- Since Lasso helps in feature reduction (as the coefficient value of some of the features become zero), Lasso has a better edge over Ridge and should be used as the final model.

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?

The 5 most important Lasso predictors are :

	Lasso Doubled Alpha Co-Efficient
RoofStyle_Shed	178633.879627
TotRmsAbvGrd	102454.772465
1stFlrSF	100211.442027
RoofMatl_WdShngl	89740.472002
RoofMatl_Metal	65024.038930
Exterior2nd_ImStucc	57453.028915
Neighborhood_NoRidge	56898.601883

Top 5 Lasso predictors after removing 5 most important predictors are:

	Lasso Co-Efficient
LotArea	115855.473197
GarageArea	84149.057562
Neighborhood_NoRidge	79401.984054
FullBath	79302.723907
HouseStyle_2.5Fin	62673.529801

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?

As Per, Occam’s Razor— given two models that show similar ’performance’ in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:- ▪ Simpler models are usually more ’generic’ and are more widely applicable ▪ Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train. ▪ Simpler models are more robust. o Complex models tend to change wildly with changes in the training data set o Simple models have low variance, high bias and complex models have low bias, high variance o Simpler models make more errors in the training set. Complex models lead to overfitting — they work very well for the training samples, fail miserably when applied to other test samples Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use. Regularization can be used to make the model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naive to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up the absolute values or the squares of the parameters of the model. Also, Making a model simple leads to Bias-Variance Trade-off: • A complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data. • A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed. Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high. Variance refers to the degree of changes in the model itself with respect to changes in the training data. Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.