

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?

Optimal value of lambda for Ridge :0.25 :::train_r2 : 0.90 ::: test_r2 :0.88

Optimal value of lambda for Lasso : 0.001 :::train_r2 : 0.89 ::: test_r2 :0.88

index	Lasso
GrLivArea	0.961592
TotalBsmtSF	0.380122
GarageArea	0.210799
CentralAir_Y	0.121911
LotArea	0.111219
Neighborhood_Somerst	0.108058
Functional_Typ	0.100399
BsmtFinSF1	0.082323
Exterior1st_BrkFace	0.075506
Neighborhood_NridgHt	0.066414
SaleType_New	0.066081
MSZoning_RL	0.062425
BsmtExposure_Gd	0.054663
KitchenAbvGr_1	0.053164

index	Ridge
GrLivArea	1.028607
TotalBsmtSF	0.382452
HouseStyle_1.5Unf	0.198357
Exterior1st_BrkFace	0.180709
HouseStyle_2.5Unf	0.171001
GarageArea	0.156667
LotArea	0.145887
Exterior1st_Stucco	0.135462
Functional_Typ	0.120893
CentralAir_Y	0.117066
Neighborhood_Somerst	0.111985
HouseStyle_SFoyer	0.101068
HouseStyle_SLvl	0.097261
BsmtExposure_Gd	0.090104

When we change the alpha value to

lambda for Ridge :0.5 :::train_r2 : 0.90 ::: test_r2 :0.88

lambda for Lasso : 0.002 :::train_r2 : 0.88 ::: test_r2 :0.88

index	Lasso
GrLivArea	0.870190
TotalBsmtSF	0.363958
GarageArea	0.247426
CentralAir_Y	0.117139
Neighborhood_Somerst	0.088188
BsmtFinSF1	0.087827
LotArea	0.086948
Functional_Typ	0.078945
SaleType_New	0.067344
MSZoning_RL	0.061340
Neighborhood_NridgHt	0.055560
Foundation_PConc	0.047870
KitchenAbvGr_1	0.046613

index	Ridge
GrLivArea	0.966165
TotalBsmtSF	0.342993
Exterior1st_BrkFace	0.172887
GarageArea	0.161584
HouseStyle_1.5Unf	0.154640
LotArea	0.148274
HouseStyle_2.5Unf	0.137778
Exterior1st_Stucco	0.126471
CentralAir_Y	0.119328
Functional_Typ	0.117993
Neighborhood_Somerst	0.109633
BsmtFinSF1	0.104250
GarageType_Attchd	0.079829
BsmtExposure_Gd	0.076975
GarageType_BuiltIn	0.074941

- There is no much difference in r2 value of both lasso and ridge regression, how ever the coefficient values is decreased when we double the alpa value

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?

- I will be choosing lasso regression since it will make some of coefficient as zero , hence remove unwanted coefficients

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?

- Below are the top variables in both ridge and lasso when we drop top 5 variables.

index	Lasso
BsmtUnfSF	0.631781
BsmtFinSF1	0.456736
LotArea	0.250027
Exterior1st_CemntBd	0.116728
GarageType_BuiltIn	0.108950
GarageFinish_RFn	0.088891
GarageFinish_Fin	0.085596
PavedDrive_Y	0.084698
GarageFinish_Unf	0.073656
Functional_Typ	0.068609

index	Ridge
BsmtUnfSF	0.688603
BsmtFinSF1	0.494247
LotArea	0.239927
CentralAir_Y	0.160604
GarageFinish_RFn	0.130549
GarageFinish_Fin	0.128034
OpenPorchSF	0.123689
GarageFinish_Unf	0.120661
GarageType_BuiltIn	0.111524
HouseStyle_2.5Unf	0.104805

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?

- Model is robust and generalisable when the complexity of model is optimisable.
- We can control the complexity by regularisation , when the alpha value is high complexity of model is decreased
- As the alpha value further increased model will become too simple, it will not be able to determined the pattrrens in data. It will underfit
- So the alpha value should be optimal to generlise the model