Housing Price Prediction – Part II

Submitter: Gnanaprakash S

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?

- The optimal value for the ridge is 10 and Lasso is 0.01.
- The R2 score for train data is 0.87 and test data is 0.90.
- There are no major changes in coefficient by doubling the alpha values.
- The most important predictors arranged by the descending in the below diagram.

Ridge Regression Model

Feature	ridge_Coefficient
GrLivArea	0.317757
OverallQual	0.211629
TotalBsmtSF	0.135944
GarageCars	0.113944
OverallCond	0.102571
Fireplaces	0.064286
Foundation_PConc	0.063974
LotArea	0.063609
2ndFlrSF	0.057454
BsmtExposure_Other	0.049948
KitchenQual_Other	0.042663
BsmtQual_Other	0.040113
MasVnrType_None	0.030760
YearRemodAdd	0.029598
Exterior1st_VinylSd	0.029013
log_MasVnrArea	0.028876
BsmtFinSF1	0.023302
RoofStyle_Other	0.017703
log_WoodDeckSF	0.016843
log_OpenPorchSF	0.011412
Foundation_Other	0.005774
LotFrontage	0.002089
LotShape_Other	0.001421
GarageFinish_Unf	0.000265
MasVnrType_Other	-0.000105
ExterQual_Other	-0.002324
BsmtQual_TA	-0.004244
LotConfig_Other	-0.004365
HouseStyle_Other	-0.008654
BedroomAbvGr	-0.015053
LotShape_Reg	-0.016536
BsmtFinType1_Unf	-0.016747
HeatingQC_Other	-0.017790
GarageType_Other	-0.019365
Exterior2nd_VinylSd	-0.031145
KitchenQual_TA	-0.031482
HouseStyle_2Story	-0.040853
ExterQual_TA	-0.058266
log_MSSubClass	-0.059918
BsmtUnfSF	-0.064516
Year_built_vs_sold	-0.125394

Feature	double_ridge_Coefficient
GrLivArea	0.305827
OverallQual	0.209156
TotalBsmtSF	0.131463
GarageCars	0.114891
OverallCond	0.099027
Fireplaces	0.065471
LotArea	0.064624
Foundation_PConc	0.062316
2ndFlrSF	0.059795
BsmtExposure_Other	0.049569
KitchenQual_Other	0.042335
BsmtQual_Other	0.040397
YearRemodAdd	0.032412
BsmtFinSF1	0.031095
MasVnrType_None	0.024162
log_MasVnrArea	0.022786
Exterior1st_VinylSd	0.021845
RoofStyle_Other	0.018867
log_WoodDeckSF	0.017596
log_OpenPorchSF	0.012310
Foundation_Other	0.005001
LotFrontage	0.002910
LotShape_Other	0.001654
MasVnrType_Other	0.000218
ExterQual_Other	-0.001240
GarageFinish_Unf	-0.001435
LotConfig_Other	-0.004401
BsmtQual_TA	-0.006212
HouseStyle_Other	-0.007311
BedroomAbvGr	-0.011634
LotShape_Reg	-0.017024
BsmtFinType1_Unf	-0.017567
HeatingQC_Other	-0.018685
GarageType_Other	-0.019783
Exterior2nd_VinylSd	-0.024398
KitchenQual_TA	-0.032844
HouseStyle_2Story	-0.037882
BsmtUnfSF	-0.055954
log_MSSubClass	-0.058554
ExterQual_TA	-0.058807
Year_built_vs_sold	-0.118584

Lasso Regression Model

Feature	lasso_Coefficient
LotFrontage	0.0
HouseStyle_2Story	0.0
RoofStyle_Other	0.0
Exterior1st_VinylSd	0.0
Exterior2nd_VinylSd	0.0
MasVnrType_None	0.0
MasVnrType_Other	0.0
ExterQual_Other	0.0
ExterQual_TA	0.0
Foundation_Other	0.0
Foundation_PConc	0.0
BsmtQual_Other	0.0
BsmtQual_TA	0.0
BsmtExposure_Other	0.0
BsmtFinType1_Unf	0.0
HeatingQC_Other	0.0
KitchenQual_Other	0.0
KitchenQual_TA	0.0
GarageType_Other	0.0
HouseStyle_Other	0.0
LotConfig_Other	0.0
LotArea	0.0
LotShape_Reg	0.0
OverallQual	0.0
OverallCond	0.0
YearRemodAdd	0.0
BsmtFinSF1	0.0
BsmtUnfSF	0.0
TotalBsmtSF	0.0
2ndFlrSF	0.0
GrLivArea	0.0
BedroomAbvGr	0.0
Fireplaces	0.0
GarageCars	0.0
log_MSSubClass	0.0
log_MasVnrArea	0.0
log_WoodDeckSF	0.0
log_OpenPorchSF	0.0
Year_built_vs_sold	0.0
LotShape_Other	0.0
GarageFinish_Unf	0.0

Feature	double_lasso_Coefficient
LotFrontage	0.0
HouseStyle_2Story	0.0
RoofStyle_Other	0.0
Exterior1st_VinylSd	0.0
Exterior2nd_VinylSd	0.0
MasVnrType_None	0.0
MasVnrType_Other	0.0
ExterQual_Other	0.0
ExterQual_TA	0.0
Foundation_Other	0.0
Foundation_PConc	0.0
BsmtQual_Other	0.0
BsmtQual_TA	0.0
BsmtExposure_Other	0.0
BsmtFinType1_Unf	0.0
HeatingQC_Other	0.0
KitchenQual_Other	0.0
KitchenQual_TA	0.0
GarageType_Other	0.0
HouseStyle_Other	0.0
LotConfig_Other	0.0
LotArea	0.0
LotShape_Reg	0.0
OverallQual	0.0
OverallCond	0.0
YearRemodAdd	0.0
BsmtFinSF1	0.0
BsmtUnfSF	0.0
TotalBsmtSF	0.0
2ndFlrSF	0.0
GrLivArea	0.0
BedroomAbvGr	0.0
Fireplaces	0.0
GarageCars	0.0
log_MSSubClass	0.0
log_MasVnrArea	0.0
log_WoodDeckSF	0.0
log_OpenPorchSF	0.0
Year_built_vs_sold	0.0
LotShape_Other	0.0
GarageFinish_Unf	0.0

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 value for the ridge is 10 and Lasso is 0.01.
- Based on analysis the Lasso coefficient value for features becomes zero and it the model is used to predict the feature selection which is important. So that Lasso has preferred than the ridge.

Question 3

After building the model, you realized 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?

Top 5 Columns

Before

Feature	ridge_Coefficient
GrLivArea	0.317757
OverallQual	0.211629
TotalBsmtSF	0.135944
GarageCars	0.113944
OverallCond	0.102571

After removing top 5

Feature	ridge_Coefficient
2ndFlrSF	0.292151
BsmtFinSF1	0.253115
BsmtUnfSF	0.201824
Fireplaces	0.142483
LotArea	0.138764

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?

- For the more robust model the balance between the training accuracy and the testing accuracy. The model should not only work with the train data well but also with the testdata.
- Simpler models are more generic and more robust.
- The model's accuracy on a testing set is a proxy for its performance in real-world applications.