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?

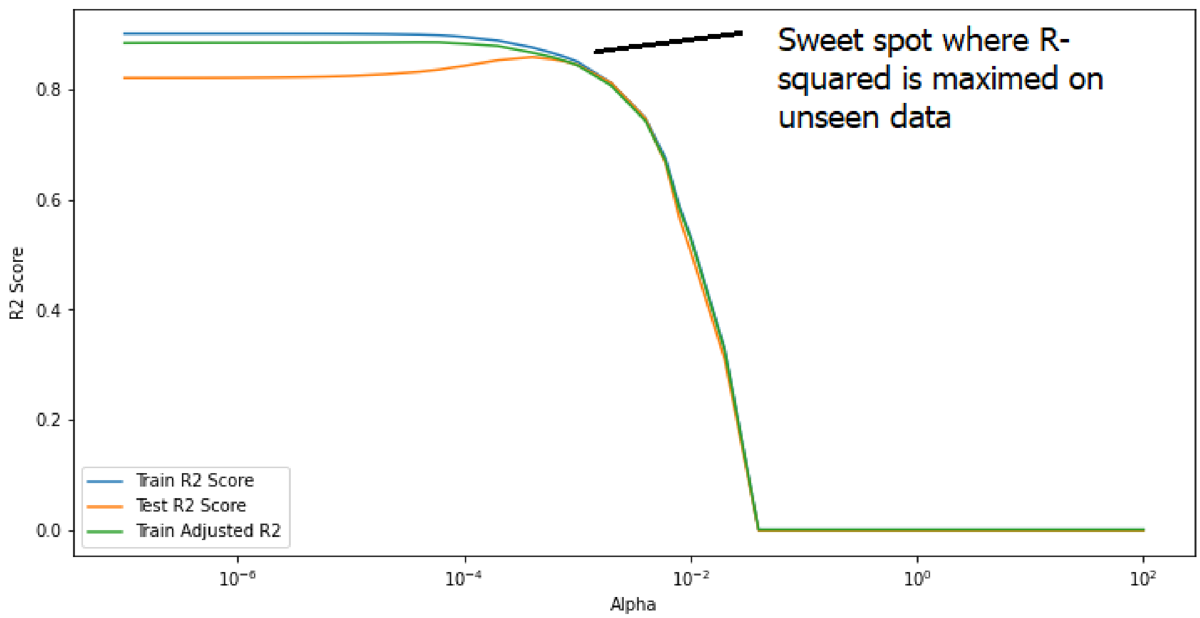
Answer:

You need to find the optimum balance between model bias and model variance when doing regularization.

The alpha regularization parameter needs to be adjusted to reach the right balance model bias and variance. The sweet spot is often found where the R-squared or adjusted R-squared values begins to taper down. You want to be slightly on the conservative side here and allow you model to reduce its accuracy somewhat so that it is more robust when handling unseen data.

Lasso and Ridge will increase regularization when the alpha regularization parameter is increased. This will reduce overfitting. Doubling the alpha parameter has a relatively small effect on the amount of regularization, but is roughly the right magnitude if you want to fine-tune the amount of regularization.

Both Lasso and Ridge will reduce the coefficients of less important predictor variables as the amount of regularization is increased. The difference is that Lasso will reduce their coefficients to zero and thereby perform feature elimination. The coefficients of the more important predictor variables will usually increase as the regularization is increased.


Sweet spot where R-squared is maximed on unseen data
R2 Score
Alpha
Train R2 Score
Test R2 Score
Train Adjusted R2
0.0
0.2
0.4
0.6
0.8
10^{-6}
10^{-4}
10^{-2}
10^{0}
10^{2}

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:

The optimal regularization values in my models were around alpha=0.0004 for Lasso and alpha=80 for Ridge. I have chosen to use a Lasso regression model for my final model because it performed feature elimination and appears to perform slightly better than the Ridge model.

Feature elimination can be useful since models with fewer features may be easier for business users to interpreted or understand. While Lasso does feature elimination automatically, one could also do feature elimination as pre-processing step for Ridge regression using RFE.

With alpha=0.0004 for regularization, the Lasso model selected 75 predictor variables out of 163. This seems to be the sweet spot, but one can increase the regularization to alpha=0.002 which results in only 29 predictors variables selected but at a very small cost in model accuracy.

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:

The top five features in my original model:

Feature	Coefficient
BsmtFullBath	0.357785
OverallCond	0.233686
LotFrontage	-0.155433
GarageArea	0.086154
Neighborhood OldTown	0.080115

The top five features in my new model where the original top five have been removed from the dataset:

Feature	Coefficient
BsmtHalfBath	0.358607
MasVnrArea	0.233213
LotArea	-0.134219
GarageQual	0.085795
Neighborhood SWISU	0.078040

Parameter	Coeficient
BsmtFullBath	0.357785
OverallCond	0.233686
LotFrontage	-0.155433
GarageArea	0.086154
Neighborhood OldTown	0.080115
Neighborhood Timber	0.066244
Neighborhood NridgHt	0.057923
Exterior CemntBd	0.050917
ExterCond	0.049402
TotRmsAbvGrd	0.041497
BsmtFinSF1	0.041328
BsmtCond	0.040899
Functional	0.039718
Neighborhood Edwards	0.039649
EnclosedPorch	0.032692
BldgType 2fmCon	0.031646
MasVnrArea	0.031629
BsmtHalfBath	0.031535
OpenPorchSF	0.030701
Exterior VinylSd	-0.028481

Figure 1 Original model

Parameter	Coeficient
BsmtHalfBath	0.358607
MasVnrArea	0.233213
LotArea	-0.134219
GarageQual	0.085795
Neighborhood SWISU	0.078040
Neighborhood Timber	0.067123
Neighborhood NridgHt	0.059114
Exterior CemntBd	0.051006
ExterCond	0.049110
TotRmsAbvGrd	0.046316
BsmtCond	0.045795
BsmtFinSF1	0.044344
Neighborhood Edwards	0.043447
Functional	0.041003
EnclosedPorch	0.038098
OpenPorchSF	0.035286
SaleType WD	0.028878
Exterior VinylSd	-0.028597
BldgType 2fmCon	0.028090
BsmtFinType2 BLQ	-0.024272

Figure 2 New model with the original top 5 features removed from the dataset

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:

Regularization is used to reduce or eliminate overfitting (high variance) of a model, but too much regularization can result in a model is that is inaccurate (high bias). You want a model that is accurate (low bias) but that is also generalizable and works well on unseen datasets (low variance).

You need to use an adequate amount of regularization on your model for it to be robust enough to handle unseen data well. Therefore, it's better to be slightly conservative and apply a bit more regularization than what seems optimal on the training dataset.