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 to 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 obtained for Ridge and Lasso in our models are –

Ridge: **5**
Lasso: **0.0001**

The top 10 predictor variables that **positively** influence the response variable are (using lasso) –

MSZoning_FV	:	0.32026
MSZoning_RM	:	0.27907
MSZoning_RL	:	0.277144
MSZoning_RH	:	0.268782
BsmtFullBath_2	:	0.201671
Condition2_PosA	:	0.167606
OverallQual_9	:	0.142449
Neighborhood_Crawfor	:	0.134168
Neighborhood_StoneBr	:	0.128639
OverallQual_8	:	0.104344

The top 10 predictor variables that **negatively** influence the response variable are (using lasso) –

Exterior1st_BrkComm	:	-0.25603
LandSlope_Sev	:	-0.25534
OverallCond_3	:	-0.2069
OverallQual_2	:	-0.15848
Functional_Maj2	:	-0.15036
Neighborhood_MeadowV	:	-0.14734
OverallCond_4	:	-0.10829
SaleCondition_Alloca	:	-0.09989
MSSubClass_160	:	-0.09298
Foundation_Wood	:	-0.08788

On doubling the coefficients i.e.

Alpha Ridge: **10**
Alpha Lasso: **0.0002**

The new top 10 predictor variables that **positively** influence the response variable are (using lasso) –

MSZoning_FV	:	0.236334
MSZoning_RL	:	0.200054
MSZoning_RM	:	0.198977
MSZoning_RH	:	0.193422
BsmtFullBath_2	:	0.162208
OverallQual_9	:	0.146284
Neighborhood_Crawfor	:	0.13214
Neighborhood_StoneBr	:	0.113624
OverallCond_9	:	0.108193
OverallQual_8	:	0.101938

The new top 10 predictor variables that **negatively** influence the response variable are (using lasso) –

OverallCond_3	:	-0.21293

Neighborhood_MeadowV	:	-0.13692
Exterior1st_BrkComm	:	-0.13179
OverallQual_2	:	-0.11235
OverallCond_4	:	-0.10151
Functional_Maj2	:	-0.10011
LandSlope_Sev	:	-0.08718
MSSubClass_160	:	-0.0831
KitchenQual_Fa	:	-0.07451
Foundation_Wood	:	-0.07052

Performance score before and after doubling the Alfa –

Alpha >>	5	0.0001	10	0.0002
Metric	Ridge Regression	Lasso Regression	Ridge Regression	Lasso Regression
R2 Score (Train)	0.942266	0.949919	0.936507	0.94532
R2 Score (Test)	0.904348	0.903411	0.905062	0.907413
RSS (Train)	7.0136	6.083891	7.7131	6.642495
RSS (Test)	5.557708	5.612136	5.516192	5.37956
MSE (Train)	0.086935	0.080969	0.008311	0.007157
MSE (Test)	0.11817	0.118747	0.013859	0.013516

Although not very significant we notice a slight increase in the error term (RSS) and a slight improvement in the test R2 score when the Alpha's are doubled. We also notice that the tolerance level improvements (MSE -Test)

We also observe that Lasso model further drops coefficient from 200 to 174 on doubling the Alpha, there by further simplifying the model.

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

In our case, we obtained the following performance scores for our Ridge and Lasso models based on the optimal value of lambda.

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.954036	0.942266	0.949919
1	R2 Score (Test)	0.867753	0.904348	0.903411
2	RSS (Train)	5.583720	7.013600	6.083891
3	RSS (Test)	7.683982	5.557708	5.612136
4	MSE (Train)	0.077569	0.086935	0.080969
5	MSE (Test)	0.138948	0.118170	0.118747

By observing the R2 scores, Ridge and Lasso seems to be performing equally well on both Training and Test Data. The difference is very minor between the R2 scores.

RSS scores on training is slightly better on Lasso than Ridge. However, the test RSS Test scores are comparable.

The MSE indicating the error tolerance are also comparable.

Since the scores for both the models are comparable, we will go ahead with the **lasso model** since it helps simplify the model complexity by removing noisy or unrelated variables with respect to our target/response variable (SalePrice).

There are 200 variables in Lasso model as compared to 291 in the Ridge model. This makes Lasso model relatively less complex to interpret as compared to Ridge 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?

Answer 3

The five most **positively** related predictor variables obtained in the lasso model are –

MSZoning_FV	:	0.32026
MSZoning_RM	:	0.27907
MSZoning_RL	:	0.277144
MSZoning_RH	:	0.268782
BsmtFullBath_2	:	0.201671

After excluding them and recreating the lasso model again, the five most **positively** related predictor variable are –

OverallQual_9	:	0.132395
Neighborhood_Crawfor	:	0.116585
SaleCondition_Partial	:	0.078388
OverallQual_8	:	0.076105
Functional_Typ	:	0.074419

The performance parameter after dropping and recreating the lasso model are –

r2_train_lr : 0.9201684252541564
r2_test_lr : 0.900028237582458
rss1_lr : 9.69803614110388
rss2_lr : 5.808697347152229
mse_train_lr : 0.010450469979637802
mse_test_lr : 0.014594716952643793

There is a slight drop in the r2 value on train dataset since we dropped some of the influencing variables. The RSS (error term) is also slightly higher on training. However, the model continues to perform well on test despite dropping the variables.

We also observe that Lasso model further drops coefficient from 200 to 88 on dropping the top five coefficient, there by further simplifying the model without underfitting (since we get a good test score)

Question 4

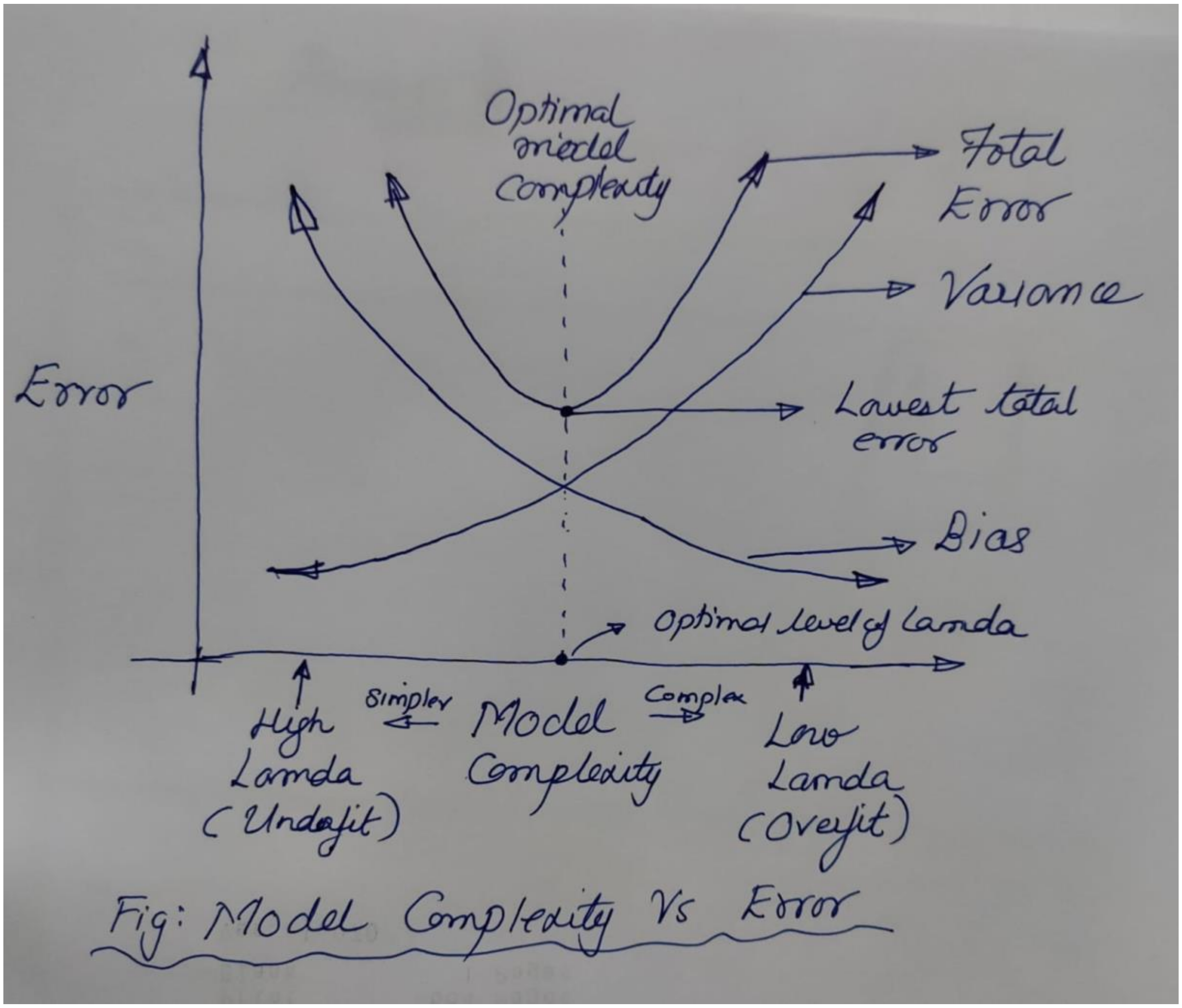
How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Answer 4

To make our model robust and generalisable we need to ensure that our model does well on both train and test (unseen data).

The model complexity is related to the magnitude of the coefficients and the number of coefficients in the model. The Regularization/Regression applied during Ridge and Lasso modelling introduces a penalty (Lambda Hyperparameter) to the cost function. This suppresses the magnitude of the coefficients towards zero or removes it entirely from the model (in case of lasso regression).

The diagram below depicts the Model complexity vs Error.



As seen from the diagram above, by introducing a small lambda(penalty), we introduce a small bias for a significant reduction in variance i.e. we manage to reduce model overfit **without losing much on the accuracy** and not compromising on the underlying patterns in the data.