Assignment Part-II

Subjective Questions

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
Here is the optimal value
- Ridge = 1.0
- Lasso = 0.0001

We see decrease in values for coefficients for Ridge and Lasso, as well as decrease in accuracy (R2 core)

Ridge Data –

	Features	Ridge wih alpha = 1	Ridge wih alpha = 2
0	MSSubClass	-0.0544	-0.0524
1	LotFrontage	-0.0276	-0.0132
2	LotArea	0.0922	0.0722
3	OverallQual	0.1770	0.1697
4	BsmtQual	0.0661	0.0631
5	BsmtExposure	0.0479	0.0486
6	TotalBsmtSF	0.0166	0.0360
7	1stFlrSF	0.1623	0.1425
8	2ndFlrSF	0.0762	0.0772
9	GrLivArea	0.1548	0.1442
10	BsmtFullBath	0.0423	0.0413
11	KitchenQual	0.0669	0.0702
12	GarageCars	0.0600	0.0634
13	Neighborhood_NoRidge	0.0791	0.0787
14	Neighborhood_NridgHt	0.0584	0.0581
15	Neighborhood_StoneBr	0.0389	0.0375
16	Exterior1st_ImStucc	-0.0264	-0.0172
17	Exterior1st_Stone	-0.0123	-0.0042
18	Exterior2nd_Stone	0.0299	0.0220
19	Foundation_Slab	0.0422	0.0409

	Metric	Ridge Regression (Lambda = 1)	Ridge Regression (Lambda = 2)
0	R2 Score (Train)	0.828541	0.826743
1	R2 Score (Test)	0.838270	0.835217
2	RSS (Train)	2.110196	2.132315
3	RSS (Test)	0.881321	0.897958
4	MSE (Train)	0.045462	0.045700
5	MSE (Test)	0.044806	0.045227

Lasso Data

	Features	Lasso wih alpha = .0001	Lasso wih alpha = .0002
0	MSSubClass	-0.0504	-0.0482
1	LotFrontage	-0.0082	-0.0000
2	LotArea	0.0807	0.0367
3	OverallQual	0.1885	0.1916
4	BsmtQual	0.0604	0.0528
5	BsmtExposure	0.0466	0.0477
6	TotalBsmtSF	-0.0000	0.0000
7	1stFlrSF	0.0696	0.0255
8	2ndFlrSF	0.0207	0.0000
9	GrLivArea	0.2842	0.3296
10	BsmtFullBath	0.0409	0.0403
11	KitchenQual	0.0651	0.0659
12	GarageCars	0.0585	0.0608
13	Neighborhood_NoRidge	0.0768	0.0748
14	Neighborhood_NridgHt	0.0563	0.0543
15	Neighborhood_StoneBr	0.0324	0.0236
16	Exterior1st_ImStucc	-0.0000	-0.0000
17	Exterior1st_Stone	-0.0000	-0.0000
18	Exterior2nd_Stone	0.0000	0.0000
19	Foundation_Slab	0.0342	0.0235

	Metric	Lasso Regression (Lambda = .0001)	Lasso Regression (Lambda = .0002)
0	R2 Score (Train)	0.827364	0.825094
1	R2 Score (Test)	0.836866	0.830456
2	RSS (Train)	2.124680	2.152616
3	RSS (Test)	0.888971	0.923902
4	MSE (Train)	0.045618	0.045917
5	MSE (Test)	0.045000	0.045875

Important predictors after Change for Ridge

1. OverallQual
2. GrLivArea
3. 1stFlrSF
4. Neighborhood_NoRidge
5. 2ndFlrSF
6. LotArea

Important predictors after Change for Lasso

1. OverallQual
2. GrLivArea
3. 1stFlrSF
4. Neighborhood_NoRidge
5. 2ndFlrSF
6. LotArea

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

I will apply Lasso as the accuracy on the Ridge & Lasso are almost similar, however with Lasso allows us feature selection and as such we can eliminate 4 features from the model, making it simpler.

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?

Answer

The five most important features now are –

1. TotalBsmtSF
2. 2ndFlrSF
3. KitchenQual
4. Foundation_Slab
5. GarageCars

Question 4

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

Answer

We can make sure that a model is robust and generalizable by making it simple so that it doesn't overfit.

Regularization can help to make the model simple by adding a penalty for the number of features added.

When we simplify the model, we comprise a little on bias, however with this trade off we reduce variance of the model, which helps in reducing error on test data.
The intention is to strike a balance between Bias and Variance for a robust and generalizable model.