Advanced Regression Assignment Subjective Questions

—By Vaibhav Parakh

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?
    a. Optimal Value of lambda for ridge: 1
    b. Optimal Value of lambda for lasso: 0.0001
    c. Observing the train and test accuracy scores of both Ridge and Lasso after doubling the value of alpha:

```
R-squared scores for Ridge Regression using Best value of alpha
best_ridge_r2_score_train 0.92229
best_ridge_r2_score_test 0.88059


R-squared scores for Ridge Regression using Double value of alpha
ridge_second_r2_score_train 0.92111
ridge_second_r2_score_test 0.88261


R-squared scores for Lasso Regression using Best value of alpha
best_lasso_r2_score_train 0.92305
best_lasso_r2_score_test 0.87777


R-squared scores for Lasso Regression using Double value of alpha
lasso_second_r2_score_train 0.92256
lasso_second_r2_score_test 0.87998
```

    d. Observing the test scores of both Ridge and Lasso after doubling the value of alpha:
        i. Double alpha Ridge model can explain 88.26% of variance of the dataset.
        ii. Double alpha Lasso model can explain 87.99% of variance of the dataset.
        iii. For Training dataset:
            1. After doubling the value of alpha for Ridge Regression, training accuracy has gone down from 92.29% to 92.11%.
            2. After doubling the value of alpha for Lasso Regression, training accuracy has gone down from 92.30% to 92.25%.
        iv. For Test dataset:
            1. After doubling the value of alpha for Ridge Regression, test accuracy has gone up from 88.05% to 88.26%.
            2. After doubling the value of alpha for Lasso Regression, test accuracy has gone up from 87.77% to 87.99%.
    e. Observing changes in the model when we choose to double the value of alpha:

i. Top 5 predictor of ridge model using optimum value of alpha (excluding intercept):

	Features	Coefficients	Absolute Coefficients
0	Intercept	-0.971649	0.971649
1	MSZoning_Residentail	0.704569	0.704569
2	Neighborhood_MeadowV	-0.504172	0.504172
3	Neighborhood_BrDale	-0.434777	0.434777
4	Neighborhood_OldTown	-0.399746	0.399746
5	GrLivArea	0.396906	0.396906

1. MSZoning_Residential has a positive relation with the SalePrice, residential plots are being sold at a higher price.
2. Neighborhood_MeadowV has a negative relation with the SalePrice, that is houses in Meadow Village are sold at a lower price.
3. Neighborood_BrDale has a negative relation with the SalePrice, that is houses in Briardale are sold at a lower price.
4. Neighborhood_OldTown has a negative relation with the SalePrice, that is houses in Old Town are sold at a lower price.
5. GrLivArea has a positive relation with the SalePrice, that is more the ground living area, higher the price the house will be sold at.

ii. Top 5 predictor of ridge model using 2x value of alpha (excluding intercept):

	Features	Coefficients	Absolute Coefficients
0	OverallQual	0.230194	0.971649
1	Intercept	-0.833609	0.971649
2	OverallCond	0.153053	0.704569
3	BsmtUnfSF	-0.139179	0.504172
4	TotalBsmtSF	0.236223	0.434777
5	GrLivArea	0.397238	0.399746

iii. Our predictors respectively for different ridge models have changed in order of influence.

iv. Top 5 predictor of lasso model using optimum value of alpha (excluding intercept):

	Features	Coefficients	Absolute Coefficients
0	Intercept	-1.164809	1.164809
1	MSZoning_Residentail	0.835011	0.835011
2	Neighborhood_Blueste	-0.610268	0.610268
3	Neighborhood_MeadowV	-0.597502	0.597502
4	Foundation_Wood	-0.500022	0.500022
5	Neighborhood_BrDale	-0.492017	0.492017

1. MSZoning_Residential has a positive relation with the SalePrice, residential plots are being sold at a higher price.
2. Neighborhood_Blueste has a negative relation with the SalePrice, that is houses in Bluestem are sold at a lower price.
3. Neighborood_MeadowV has a negative relation with the SalePrice, that is houses in Meadow Village are sold at a lower price.
4. Foundation_wood has a negative relation with the SalePrice, that is house with a wooden foundation is sold at a lower
5. Neighborhood_BrDale has a negative relation with the SalePrice, that is houses in Briardale are sold at a lower price.

v. Top 5 predictor of lasso model using 2x value of alpha (excluding intercept):

	Features	Coefficients	Absolute Coefficients
0	Intercept	-1.055809	1.055809
1	MSZoning_Residentail	0.821251	0.821251
2	Neighborhood_MeadowV	-0.567746	0.567746
3	Neighborhood_Blueste	-0.509847	0.509847
4	Neighborhood_BrDale	-0.472582	0.472582
5	Foundation_Wood	-0.466930	0.466930

vi. Our predictors respectively have changed for different lasso models in order of influence.

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?
   a. Optimal Value of lambda for ridge: 1
   b. Optimal Value of lambda for lasso: 0.0001

```
R-squared scores for Ridge Regression using best value of alpha
best_ridge_r2_score_train 0.922
best_ridge_r2_score_test 0.881


R-squared scores for Lasso Regression using best value of alpha
best_lasso_r2_score_train 0.923
best_lasso_r2_score_test 0.878
```

   c. Observing the test scores of both Ridge and Lasso:
      i. Best Ridge model can explain 88.1% of variance of the test dataset.
      ii. Best Lasso model can explain 87.8% of variance of the test dataset.

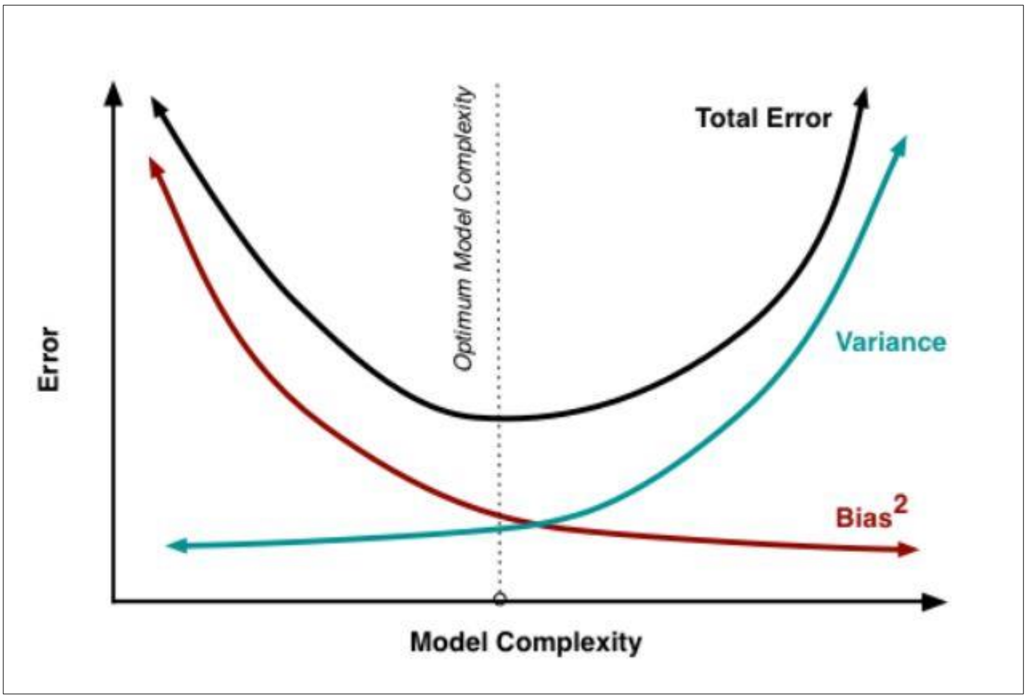
d. Given how close the model performance is, I would go with lasso since it also has inbuilt feature selection (by reducing coefficients of junk features to 0) which would make the model little more robust than ridge and hence would have a higher probability of a better performance on more unseen data.

3. After building the model, you realized that the first 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?

   As per the image below, here are the top 5 predictors.

	Features	Coefficients	Absolute Coefficients
0	Exterior1st_BrkFace	0.554926	0.554926
1	Exterior1st_WdShing	0.519290	0.519290
2	Intercept	-0.503421	0.503421
3	BldgType_Twnhs	-0.441643	0.441643
4	GrLivArea	0.400445	0.400445
5	RoofStyle_Shed	0.399199	0.399199

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?



As per the above diagram, there is a tradeoff between Bias-Variance as we keep on increasing the model complexity, that is at extremely high model complexity, bias is at its lowest and variance is at its peak. What this implies is that, on training set, because the model is so complex, it can learn all the data points, identify all the underlying patterns to an extent that it can predict with tending to 100% accuracy. But, as soon as the model is tested on the unseen portion of data (test data set for example) we get to see a dismally low accuracy score which indicates that the model perhaps has an extremely high variance, that is, as soon as unseen dataset is introduced, the model is unable to identify underlying patterns basis the underlying patterns of the train dataset.

To deal with the above scenario, i.e., overfitting on the train dataset and poor fitting on the test dataset, we aim to build a model that is robust enough that it can identify patterns of both train and dataset with near about similar accuracy in both the cases.

To make sure the model is generalizable, we can add a regularization component, i.e., either ridge or lasso. In case we go with the decision of going with ridge regularization, higher the value of alpha (lambda) more is the probability of shrinking the coefficients to close to 0 thus ensuring junk coefficients do not have an impact on the model thereby ensuring variance is low while having some more margin of error (not completely overfitting the data).

If we go with lasso, at an extremely high value of alpha (lambda) it will shrink junk coefficients to 0 to ensure only features which have significant impact on the model are present as part of the model.

Lasso, in essence, selects prominent features and is preferred more than ridge especially when there are a few features which have more influence on the prediction. If all the features have the same influence on the target variable (y) we go with ridge regression.