

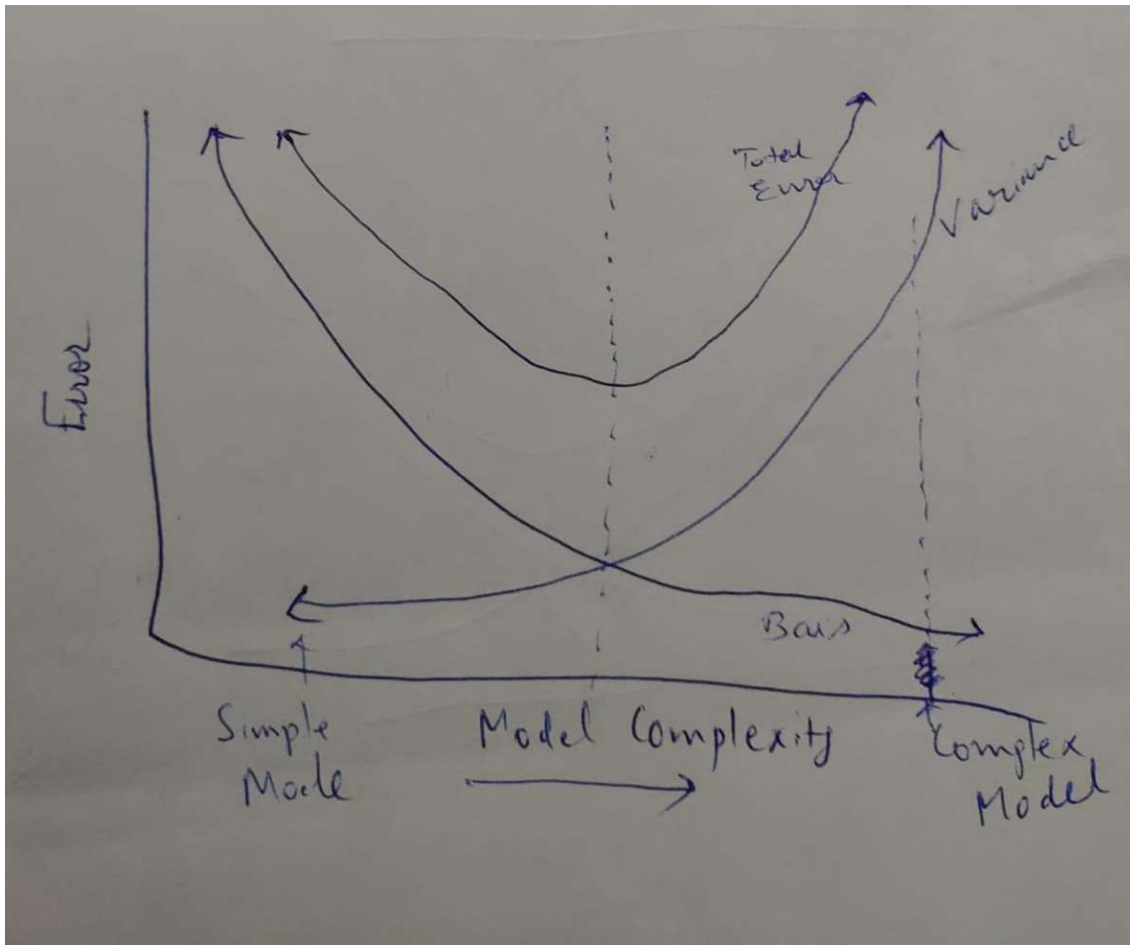
Assignment-based Subjective Questions

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 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:

The optimal value of alpha for Ridge and Lasso regression is a value where small compromise in bias, will get significant reduction in variance. As per assignment model, the optimal value of Alpha for Ridge and Lasso regression are 4.0 and .0001 respectively. If we double the value of alpha, then model will become more biased and underfitting. "GrLivArea" is most important predictor variable as regularized model.

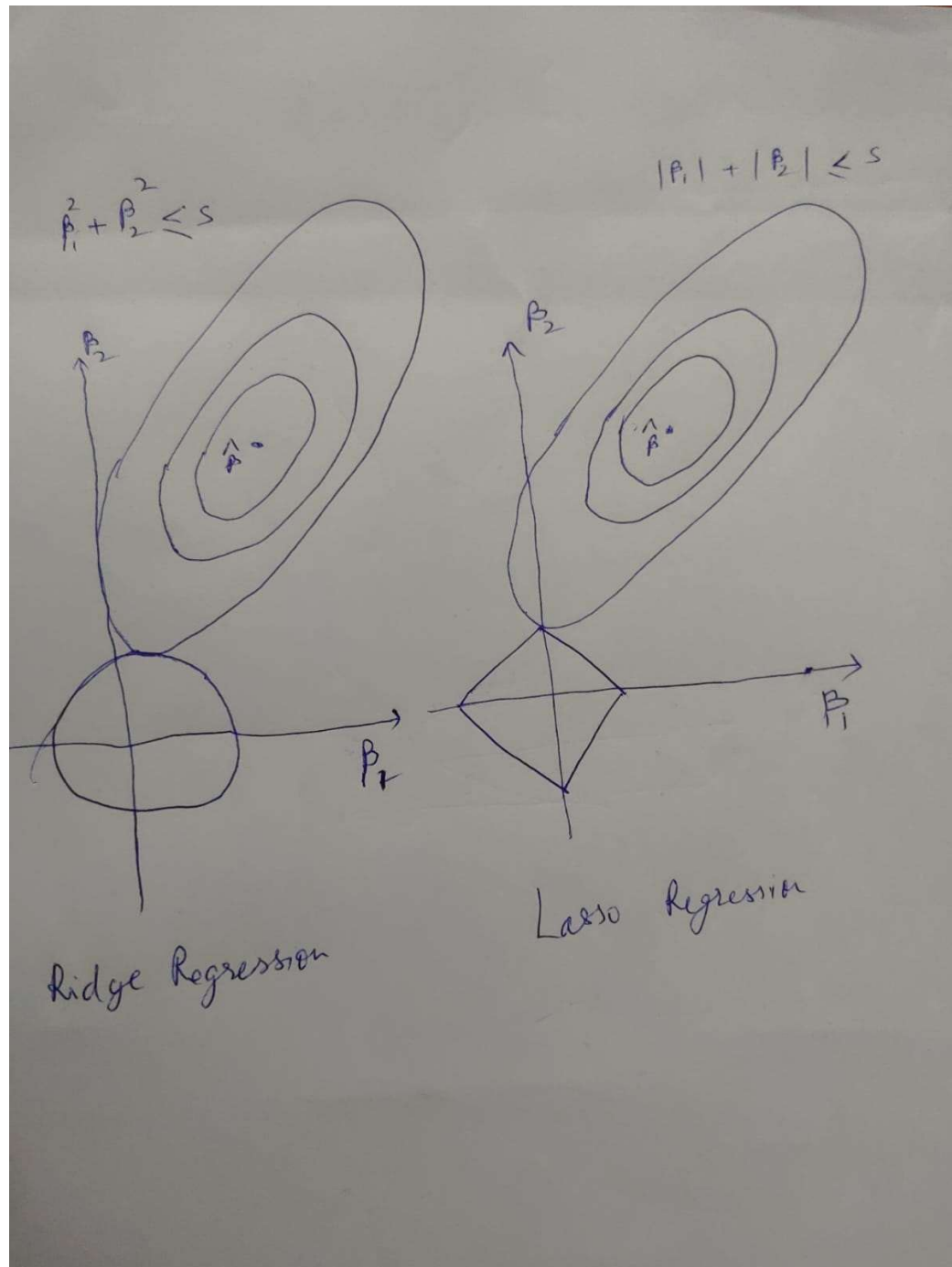


Question 2

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:

I will recommend using "Lasso regression" model because it is not only reducing variances, but also do feature selection which reduces noise in prediction. R2 Score (Test) is better for Lasso regression than Ridge regression. i.e. 0.880322



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:

Following are top 5 predictor variables on basis of Lasso regression model –

1. GrLivArea (Above grade (ground) living area square feet)
2. TotalBsmtSF (Total square feet of basement area)
3. OverallQual (Rates the overall material and finish of the house)
4. SaleType (Type of sale)
5. YearBuilt (Original construction date)

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:

We will refer different metrics like R2 Score, RSS & TMS for robust & generalisable of model. If R2 Score on test is high, model is generalizable. If R2 Score on train is high, it is robust. R2 score also show that how accurately model predictions are. If R2 score is higher, then it more accurately predicts.

Coefficient of determination (R_2)	Interpretation
0	The model <i>does not predict</i> the outcome.
Between 0 and 1	The model <i>partially predicts</i> the outcome.
1	The model <i>perfectly predicts</i> the outcome.