Subjective Answers

Q1 . 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?

Ans.1

The optimal lambda value for Ridge and lasso regression is shown below:

- Ridge - 50
- Lasso - 0.0001

Below are the top 5 predictors for Lasso after doubling the value of alpha:

Features	rfe_support	rfe_ranking	Coefficient	
0	MSZoning_RM	True	1	0.085771
1	OverallQual	True	1	0.074976
2	MSZoning_FV	True	1	0.073226
3	TotalBsmtSF	True	1	0.058045
4	Foundation_PConc	True	1	0.042136

Below are the top 5 predictors for Ridge after doubling the value of alpha:

Features	rfe_support	rfe_ranking	Coefficient	
0	OverallQual	True	1	0.074976
6	GrLivArea	True	1	0.072381
2	TotalBsmtSF	True	1	0.058045
1	BsmtFinSF1	True	1	0.049500
4	2ndFlrSF	True	1	0.046515

Below are the changes observed in terms of accuracy, mean squared error, RSS of train and test set.

	Ridge	Ridge1(alpha double)	Lasso	Lasso1(alpha double)
R^2 for Train set(Acurracy)	0.9090214796338847	0.9069671867004614	0.9119587677725383	0.9119112110449746
R^2 for Test test (Acurracy)	0.8942986596420384	0.8931810426954747	0.8988147542979065	0.8987128211216182
RSS for Train test	12.160253741358893	12.434831995973065	11.767653719576012	11.774010185607283
RSS for Test test	5.999833426588248	6.063271747156419	5.743490265922991	5.749276210323521
MSE for Train test	0.012800267096167257	0.013089296837866384	0.012387003915343172	0.012393694932218192
MSE for Test test	0.014705474084775117	0.014860960164599066	0.014077182024321056	0.014091363260596866

Q2. 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?

Ans.1 We will choose Lasso regression because the mean squared error of lasso is less than Ridge regression.

The optimal lambda value for Ridge and lasso regression is shown below:

- Ridge - 50
- Lasso - 0.0001

The Mean Squared error for Ridge and Lasso :

- Ridge - 0.0147054
- Lasso - 0.0140771

The Mean Squared Error of Lasso is slightly lower than that of Ridge. Also Lasso shrinks the coefficients of features to exactly 0. Lasso helps in feature reduction. Lasso has a better edge over Ridge.

Therefore , the variables predicted by Lasso can be applied to choose significant variables for predicting the price of a house.

Q3. 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?

Ans3 . Below are the top 5 predictors when rebuilding the model after removing five most predictor variables in the lasso.

	Features	rfe_support	rfe_ranking	Coefficient
5	2ndFlrSF	True	1	0.133481
4	1stFlrSF	True	1	0.097295
3	TotalBsmtSF	True	1	0.073139
2	BsmtFinSF1	True	1	0.053868
12	Foundation_PConc	True	1	0.052063

Q4. 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?

Ans.

Per , Occam's Razor – given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:

- Simpler models, are usually more 'generic' and are more widely applicable.
- Simpler models require fewer training samples for effective training than the complex ones and hence are easier to train.

- Simpler models are more robust.
  - Complex models tend to change with changes in the training dataset.
  - Simple models have low variance, high bias whereas complex models have low bias and high variance.

- Simpler models make more errors in the train set. Complex models lead to overfitting but fail miserably when applied to other test samples.
- Therefore, to make the model more robust and generalizable, make the model simple but not simpler because then it will not be any use.

  Regularization can be used to solve the problem of overfitting.It helps with managing model complexity by essentially shrinking the model coefficient estimates towards 0. This discourages the model from becoming too complex , thus avoiding the risk of overfitting. It also solves the multicollinearity issue.
  Cost function = RSS + Penalty

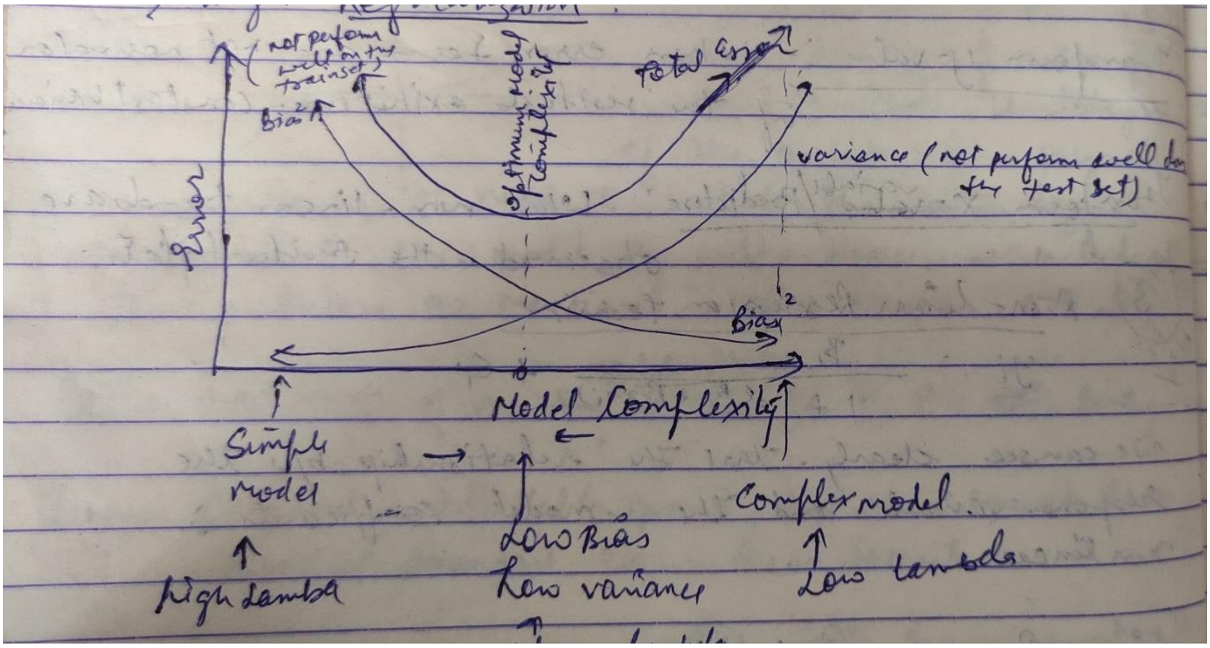
  Here regularization involves adding a regularization term to the cost that adds up the absolute or the squares of the Beta parameters (coeff.) of the models.

  Also, making the model simple leads to Bias-Variance Trade-Off.

- A Complex model will need to change for every little change in the train set. Therefore, its unstable and highly sensitive to any changes in the train set.
- A simpler model that finds out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

  Bias quantifies how accurate is the model likely to be on test data.

  Variance refers to the degree of changes in the model itself with respect to change in the train set.



(not perform well on the trainset)
Bias²
optimum model complexity
Total error
variance (not perform well on the test set)
Error
Bias²
Model Complexity
Simple model
Low Bias
Low variance
Complex model
High lamba
Low lambda