Subjective Questions:

Please note all the subjective questions are worked out on the section Part2 of python notebook which is submitted.

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?

Q1-Part1:

What is the optimal value of alpha for ridge and lasso regression?

- Optimal Alpha for Ridge - 8
- Optimal Alpha for Lasso - 0.001

Q1-Part2:

What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

There is a slight change in the train and test scores after doubling of alpha but in general there is no major impact to the model other than few changes in co-efficient of predictor variables

Q1-Part3:

What will be the most important predictor variables after the change is implemented?

Ridge with Best Alpha - 8

- Neighborhood_Crawfor 0.119365
- GrLivArea 0.111425
- Functional_Typ 0.094283
- Exterior1st_BrkFace 0.082414
- Neighborhood_StoneBr 0.065533

Ridge with Double Alpha - 16

- GrLivArea 0.108544
- Neighborhood_Crawfor 0.094699
- Functional_Typ 0.082434
- Exterior1st_BrkFace 0.064683
- TotalBsmtSF 0.056722

Lasso with Best Alpha - 0.001

- GrLivArea 0.135688
- Neighborhood_Crawfor 0.116623
- Functional_Typ 0.114918
- Neighborhood_Somerst 0.081844
- SaleType_New 0.076526

Lasso Double Alpha - 0.002

- GrLivArea 0.140115
- Functional_Typ 0.104581
- Neighborhood_Crawfor 0.081813
- TotalBsmtSF 0.062239
- SaleType_New 0.052753

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?

	Metric	Linear Regression	Ridge Regression	Ridge+RFE Regression	Lasso Regression
0	R2 Score (Train)	9.442843e-01	0.934489	0.892108	0.918193
1	R2 Score (Test)	-8.597513e+21	0.880799	0.851519	0.875330
2	RSS (Train)	7.876221e+00	9.260922	15.252099	11.564663
3	RSS (Test)	5.561424e+23	7.710660	9.604703	8.064425
4	MSE (Train)	8.942127e-02	0.096964	0.124436	0.108355
5	MSE (Test)	3.625958e+10	0.135013	0.150686	0.138075

Observation:

- Based on the above table Lasso can be selected as final model since it has the least variation between **train score(0.92) and test score(0.88)** while having above 90% train score.

<u>Question 3</u>

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?

Top 5 predictors for Lasso

- GrLivArea
- Functional
- Neighborhood
- SaleType
- TotalBsmtSF

After removing top 5 predictors below are the new top predictors

- BsmtFinSF1 0.155892 - Original Variable(**BsmtFinSF1**)
- BsmtUnfSF 0.130769 - Original Variable(**BsmtUnfSF**)
- Exterior1st_BrkFace 0.100354 - Original Variable(**Exterior1st**)
- 2ndFlrSF 0.099956 - Original Variable(**2ndFlrSF**)
- CentralAir_Y 0.064463 - Original Variable(**CentralAir**)

Question 4

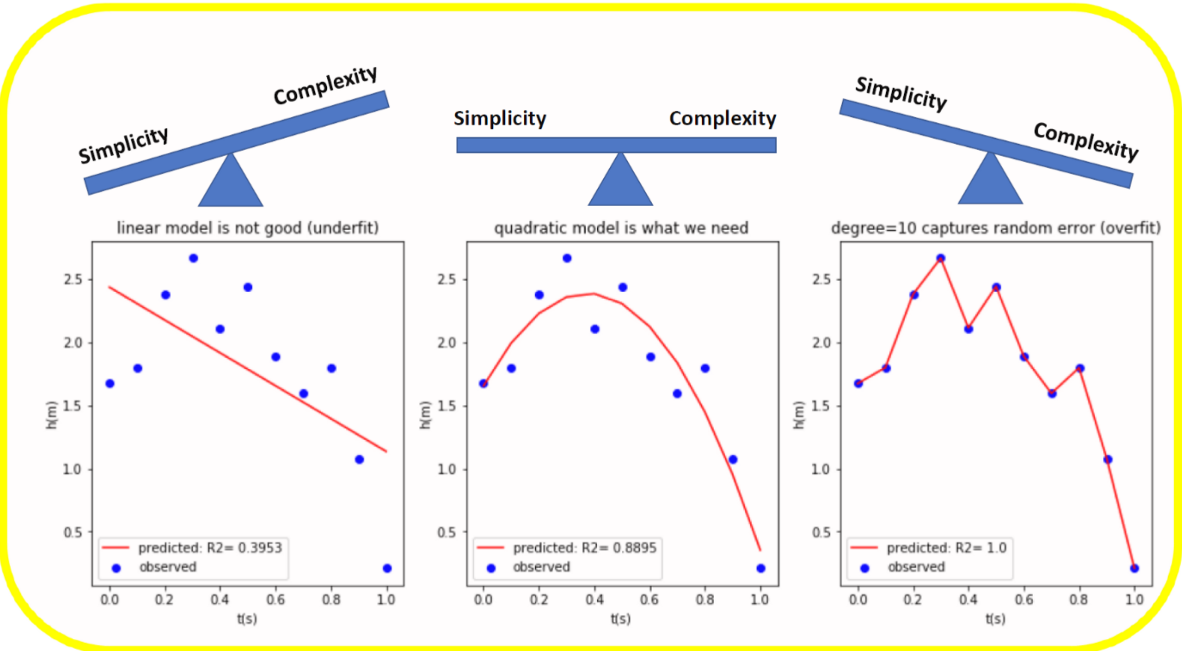
How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

Model selection is the process of choosing one from among possibly many candidate machine learning models for a predictive modeling project.

It is often straightforward to select a model based on its expected performance, e.g. choose the model with the highest accuracy or lowest prediction error.

Another important consideration is to choose simpler models over complex models.

Simpler models are typically defined as models that make fewer assumptions or have fewer elements, most commonly characterized as fewer coefficients (e.g. rules, layers, weights, etc.). The rationale for choosing simpler models is tied back to Occam's Razor.



In statistics and machine learning, the bias-variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples and vice versa. The bias-variance dilemma or problem is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised learning algorithms from generalizing beyond their training set:

The bias is an error from erroneous assumptions in the learning algorithm. High bias (overly simple) can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).

The variance is an error from sensitivity to small fluctuations in the training set. High variance (overly complex) can cause an algorithm to model the random noise in the training data, rather than the intended outputs (overfitting).

3 Reasons why a simple model is preferred over a complex model

Prevents Overfitting: A high-dimensional dataset having too many features can sometimes lead to overfitting (model captures both real and random effects).

Interpretability: An over-complex model having too many features can be hard to interpret especially when features are correlated with each other.

Computational Efficiency: A model trained on a lower-dimensional dataset is computationally efficient (execution of algorithm requires less computational time).

Regularization is one of the most important concepts of machine learning. It is a technique to prevent the model from overfitting by adding extra information to it.

Sometimes the machine learning model performs well with the training data but does not perform well with the test data. It means the model is not able to predict the output when deals with unseen data by introducing noise in the output, and hence the model is called overfitted. This problem can be deal with the help of a regularization technique.

This technique can be used in such a way that it will allow to maintain all variables or features in the model by reducing the magnitude of the variables. Hence, it maintains accuracy as well as a generalization of the model.

It mainly regularizes or reduces the coefficient of features toward zero. In simple words, "In regularization technique, we reduce the magnitude of the features by keeping the same number of features."