Question 1

a) What is the optimal value of alpha for ridge and lasso regression?
b) What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?
c) What will be the most important predictor variables after the change is implemented?

Answer 1

a) optimal value for ridge model = {'alpha': 2.0}
optimal value for lasso model = {'alpha': 0.0001}

b)
For ridge, if double the alpha value to 4, we see following changes in the performance metrics:

Metric (for Ridge)	Optimal alpha	Double alpha
r2 score of training set	0.9308	0.9289
r2 score of testing set	0.8771	0.8786
rss value of training set	64.1261	65.9847
rss value of testing set	48.9194	48.3069
MSE value of training set	0.0691	0.0711
MSE value of testing set	0.1229	0.1214

Observations:
1. r2 score for train data reduced by approximately 0.002 while it increased for test data by 0.001 approximately
2. rss increased for test data by 1.8 but reduced for test data by 0.6 approximately
3. MSE increased for test data by 0.002 but reduced for test data by 0.0016 approximately

For lasso, if double the alpha value to 0.002, we see following change in the performance metrics:

Metric (for Lasso)	Optimal alpha	Double alpha
r2 score of training set	0.9328	0.9322
r2 score of testing set	0.8727	0.8749
rss value of training set	62.3581	62.9417
rss value of testing set	50.6586	49.7669
MSE value of training set	0.0672	0.0678
MSE value of testing set	0.1273	0.1250

Observations:
1. r2 score for train data reduced by approximately 0.0007 while it increased for test data by 0.0022 approximately
2. rss increased for test data by 0.59 but reduced for test data by 0.89 approximately
3. MSE increased for test data by 0.0007 but reduced for test data by 0.002 approximately

c)

For Ridge model if we double the alpha value to 4, we see following changes in top 20 important features

Ridge Feaure	Ridge Coef	Ridge Feaure	Ridge Coef (for double alpha)
BsmtCond_2.0	0.360527	BsmtCond_2.0	0.354604
SaleType_Oth	0.355191	SaleType_Oth	0.351172
Neighborhood_NridgHt	0.345379	Neighborhood_NridgHt	0.326954
GarageArea	0.322681	GarageArea	0.320854
FireplaceQu_5.0	0.290445	FireplaceQu_5.0	0.279091
BsmtFinType1_GLQ	0.284677	BsmtFinType1_GLQ	0.275258
BsmtFinType2_4.0	0.276520	BsmtFinType2_4.0	0.247357
MSZoning_RL	0.245554	2ndFlrSF	0.241600
Neighborhood_OldTown	0.239596	Neighborhood_OldTown	0.237931
2ndFlrSF	0.239565	MSZoning_RL	0.231263
Foundation_Stone	0.232590	Foundation_Stone	0.222319
GarageType_CarPort	0.210526	GarageType_CarPort	0.180418
SaleType_ConLD	0.192483	OverallCond	0.173803
ExterQual_4	0.174209	Neighborhood_Timber	0.149061
Neighborhood_Timber	0.166886	ExterQual_4	0.148813
OverallCond	0.162276	Condition1_PosA	0.142968
Condition1_RRAe	0.153710	1stFlrSF	0.141179
Condition1_PosA	0.145752	SaleType_ConLD	0.137503
1stFlrSF	0.143538	Condition1_RRAe	0.129207
GarageFinish_3.0	0.138511	LotConfig_FR3	0.126964

Top features for Ridge model:

When we double the alpha from optimal values, we observe following:

- Top 7 features remain the same
- 8th and 10^{th} feature got swapped, wile 9^{th} remained the same
- What is interesting is 19 out of 20 features are same in both(at different rank), onl the 20^{th} feature in both is different
- the most important feature, 'BsmtCond_2.0', is same in both. However its coefficient value is reduced from 0.36 to 0.35 , making it a little less significant

For Lasso model if we double the alpha value to 0.002, we see following changes in top 20 important features:

Lasso Feaure	Lasso Coef
BsmtCond_2.0	0.366245
SaleType_Oth	0.345030
Neighborhood_NridgHt	0.330837
GarageType_CarPort	0.329764
GarageFinish_3.0	0.326146
GarageArea	0.325887
FireplaceQu_5.0	0.305693
BsmtFinType2_4.0	0.304968
BsmtFinType1_GLQ	0.292241
SaleType_ConLD	0.289436
Exterior2nd_Brk Cmn	0.258973
Foundation_Stone	0.239510
2ndFlrSF	0.236076
GarageType_Basment	0.224828
Neighborhood_OldTown	0.214761
MSZoning_RL	0.214557
GarageType_Detchd	0.212759
GarageType_None	0.209332
ExterQual_4	0.200920
SaleCondition_Normal	0.189102

Lasso Feaure	Lasso Coef (for double alpha)
BsmtCond_2.0	0.368186
Neighborhood_NridgHt	0.356684
SaleType_Oth	0.352525
GarageArea	0.323802
FireplaceQu_5.0	0.302482
BsmtFinType2_4.0	0.298335
BsmtFinType1_GLQ	0.291491
GarageType_CarPort	0.265161
SaleType_ConLD	0.260242
2ndFlrSF	0.237516
Foundation_Stone	0.235846
Neighborhood_OldTown	0.235620
MSZoning_RL	0.231007
GarageFinish_3.0	0.220786
Neighborhood_Timber	0.176893
Condition1_RRAe	0.171307
ExterQual_4	0.171095
GarageType_Basment	0.157821
1stFlrSF	0.156762
OverallCond	0.150877

Top features for Lasso model:
While the most important feature is still BsmtCond_2. 0 (with a little higher coefficient value), there is some changes in other top features when we double the alpha from optimal value:

- Most important feature BsmtCond_2.0 coefficient value has increased a little from 0.366245 to 0.368186
- Neighborhood_NridgHt which was 3rd from top as now gone up to 2nd position
- SaleType_Oth moved down, from 2nd to 3rd position
- GarageType_CarPort and GarageFinish_3.0 which were on 4th and 5th position, is now replaced by GarageArea and FireplaceQu_5.0
- 14 out of 20 features are same in both(at different rank), but 6 new features have replaced the original selection

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 2
Ridge regression and Lasso regression are two popular techniques that make use of regularization for predicting.

Both the techniques work by penalizing the magnitude of coefficients of features along with minimizing the error between predictions and actual values or records.

The key difference however, between Ridge and Lasso regression is that Lasso Regression has the ability to nullify the impact of an irrelevant feature in the data, by reducing the coefficient of a feature to zero thus completely eliminating it and hence is better at reducing the variance when the data consists of many insignificant features.

Ridge regression, on the other hand, cannot reduce the coefficients to absolute zero. However, it performs better when the data consists of features which are sure to be more relevant and useful.

Mathematically,
Lasso is = Residual Sum of Squares + λ * (Sum of the absolute value of the magnitude of coefficients)
whereas
Ridge is = Residual Sum of Squares + λ * (Sum of the square value of the magnitude of coefficients)

Which makes Ridge mathematically more expensive.

I will go for Lasso regression in spite of the fact that the metric values are not very different for Ridge and Lasso (and even simple LR) for the following reasons:

1. We have a huge bank of features and there are many features that does not look that strongly related to target (evident from heatmap and correlation values). This is exactly the scenario where Lasso is useful.

If we look at the coefficients values for simple LR, Ridge and Lasso, what we will notice is Lasso has **managed to reduce 7 variables' coefficients to zero and hence managed to eliminate 7 features** that is irrelevant, namely - BsmtUnfSF , Exterior1st_AsphShn , Exterior2nd_Brk Cmn, HeatingQC_1 , HeatingQC_3, GarageQual_2.0 , GarageQual_5.0

This shows that Lasso managed to get same performance with less number of features by eliminating irrelevant features and hence it is a simpler model, with lesset features giving same performance on given training / testing data.

When we have two (or three) competing models that fit the data equally well, Occam's razor recommends to 'shave away all but what is necessary'. The concept of parsimony is based on Occam's razor, which also proposes that the model with fewer parameters to be preferred to the one with more. Which means Lasso is the best option out of the three models.

While Ridge and simple LR retained those variables may be with smaller coefficient values.

2. Mathematically, Lasso is less expensive as mentioned earlier

Coefficient Values for all three model optimal alpha (For reference):

	Simpple Linear	Ridge	Lasso
OverallQual	1.390292e-01	0.162276	0.145833
OverallCond	8.184353e-02	0.085248	0.084518
BsmtFinSF1	1.108733e-01	0.109412	0.095934
BsmtFinSF2	2.622608e-02	0.026510	0.021605
BsmtUnfSF	1.590126e-02	0.011330	-0.000000
TotalBsmtSF	1.502651e-01	0.143538	0.159286
1stFlrSF	2.330845e-01	0.239565	0.236076
2ndFlrSF	3.286632e-01	0.322681	0.325887
GarageArea	8.810272e-02	0.084911	0.084915
MSZoning_FV	1.988933e-01	0.245554	0.214557
MSZoning_RL	1.129495e-01	0.104459	0.100121
MSZoning_RM	1.255535e-01	0.035636	0.092625
LotConfig_CulDSac	1.236575e-01	0.127090	0.124135
LotConfig_FR3	-1.880254e-01	-0.115360	-0.163729
Neighborhood_Blueste	-3.977103e-01	-0.082218	-0.251812
Neighborhood_BrDale	-2.744639e-01	-0.098075	-0.209973
Neighborhood_BrkSide	-3.744250e-01	-0.184115	-0.316937
Neighborhood_ClearCr	-1.272219e-01	-0.011965	-0.080181
Neighborhood_CollgCr	-1.646902e-01	-0.065226	-0.127036
Neighborhood_Edwards	-4.393672e-01	-0.302150	-0.395429
Neighborhood_Gilbert	-1.598933e-01	-0.062423	-0.122521
Neighborhood_IDOTRR	-4.653567e-01	-0.269180	-0.417414
Neighborhood_MeadowV	-5.389210e-01	-0.270293	-0.459040
Neighborhood_Mitchel	-4.129433e-01	-0.283978	-0.372407
Neighborhood_NAmes	-3.952470e-01	-0.270850	-0.352570
Neighborhood_NPkVill	-2.100645e-01	-0.000033	-0.056290
Neighborhood_NWAmes	-2.945815e-01	-0.175136	-0.250728
Neighborhood_NoRidge	2.995924e-01	0.345379	0.330837
Neighborhood_NridgHt	1.892830e-01	0.239596	0.214761
Neighborhood_OldTown	-5.548004e-01	-0.363763	-0.500771
Neighborhood_SWISU	-4.499224e-01	-0.292533	-0.406272
Neighborhood_Sawyer	-3.429288e-01	-0.215161	-0.301932
Neighborhood_SawyerW	-1.461155e-01	-0.048502	-0.107311

	Simpple Linear	Ridge	Lasso
Neighborhood_StoneBr	1.332890e-01	0.166886	0.157916
Neighborhood_Timber	-1.794557e-01	-0.099804	-0.148923
Neighborhood_Veenker	-1.269716e-01	-0.023993	-0.078225
Condition1_Norm	1.500474e-01	0.145752	0.148601
Condition1_PosA	1.733587e-01	0.096588	0.141323
Condition1_PosN	1.930546e-01	0.153710	0.181493
Condition1_RRAe	-1.888409e-01	-0.145732	-0.174421
BldgType_Duplex	-2.992230e-01	-0.254322	-0.291763
BldgType_Twnhs	-3.891245e-01	-0.321611	-0.379012
BldgType_TwnhsE	-2.349453e-01	-0.193178	-0.228241
HouseStyle_1.5Unf	4.494721e-02	0.007129	0.003709
HouseStyle_2.5Fin	-4.829215e-01	-0.259469	-0.432690
HouseStyle_2.5Unf	-1.557614e-01	-0.081346	-0.119775
HouseStyle_SFoyer	2.744041e-02	0.002704	0.016297
RoofStyle_Gambrel	-1.617140e-01	-0.094895	-0.130280
Exterior1st_AsphShn	-1.741094e-15	0.000000	0.000000
Exterior1st_BrkComm	-7.145946e-01	-0.160624	-0.477571
Exterior1st_CemntBd	1.157881e-01	0.097728	0.111876
Exterior2nd_AsphShn	3.912505e-01	0.105227	0.258973
Exterior2nd_Brk Cmn	1.356401e-01	-0.092328	-0.000000
MasVnrType_BrkFace	1.292180e-01	0.068002	0.095833
MasVnrType_None	1.082791e-01	0.045547	0.073957
MasVnrType_Stone	2.308645e-01	0.174209	0.200920
ExterQual_4	2.440694e-01	0.232590	0.239510
Foundation_Stone	-4.450522e-01	-0.222221	-0.391309
Foundation_Wood	-2.426066e-01	-0.128843	-0.195431
BsmtQual_5.0	3.626182e-01	0.360527	0.366245
BsmtCond_2.0	-1.573739e-01	-0.146969	-0.151464
BsmtExposure_4.0	2.930177e-01	0.284677	0.292241
BsmtFinType1_GLQ	1.077462e-01	0.108935	0.108172
BsmtFinType1_LwQ	-1.615473e-01	-0.146218	-0.155463
BsmtFinType1_None	3.149193e-01	0.276520	0.304968
BsmtFinType2_4.0	-1.519622e-01	-0.161984	-0.159820
BsmtFinType2_6.0	1.535413e-01	0.115151	0.143695

	Simpple Linear	Ridge	Lasso
HeatingQC_1	-2.078494e-01	-0.011053	0.000000
HeatingQC_2	-2.445657e-01	-0.038893	-0.027224
HeatingQC_3	-2.152468e-01	-0.011432	-0.000000
HeatingQC_4	-1.588872e-01	0.061379	0.062894
KitchenQual_4	3.104444e-01	0.290445	0.305693
FireplaceQu_5.0	1.451651e-01	0.130942	0.138123
GarageType_Attchd	3.624809e-01	0.103600	0.224828
GarageType_Basment	1.992805e-01	-0.060479	0.053409
GarageType_BuiltIn	4.654808e-01	0.210526	0.329764
GarageType_CarPort	3.852460e-01	0.067821	0.212759
GarageType_Detchd	3.479758e-01	0.078695	0.209332
GarageType_None	7.345917e-01	0.138511	0.326146
GarageFinish_3.0	8.219602e-02	0.088363	0.083764
GarageQual_2.0	2.616220e-01	-0.034269	0.000000
GarageQual_3.0	3.173641e-01	0.009195	0.053470
GarageQual_4.0	3.894104e-01	0.071750	0.113037
GarageQual_5.0	3.502734e-01	0.006438	0.000000
SaleType_CWD	3.203123e-01	0.192483	0.289436
SaleType_ConLD	-2.583660e-01	-0.144179	-0.217685
SaleType_ConLw	1.451141e-01	0.102186	0.123751
SaleType_New	3.323093e-01	0.355191	0.345030
SaleType_Oth	2.217571e-01	0.128782	0.189102
SaleCondition_Normal	1.263849e-01	0.136191	0.133651

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 3

The most important 5 features in previous Lasso model with optimal alpha and all features were as follows (from higher to lower importance)

1. BsmtCond : Evaluates the general condition of the basement
   BsmtCond_2.0 to be precise: Fa Fair - dampness or some cracking or settling
2. SaleType : Type of sale
   SaleType_Oth to be precise wich means other
3. Neighborhood : Physical locations within Ames city limits
   Neighborhood_NridgHt to be precise which is Northridge Heights area, a premium locality
4. GarageType : Garage location , GarageType_CarPort to be precise which means Car Port
5. GarageFinish : Interior finish of the garage
   GarageFinish_3.0 to be precise which means Finished

NEXT 5 Features were:
GarageArea
FireplaceQu_5.0
BsmtFinType2_4.0
BsmtFinType1_GLQ
SaleType_ConLD

The Most important 5 features in new Lasso model are as follows (from higher to lower importance)

- BsmtFinType2_4.0 : Basement Finish Type TA Typical - slight dampness allowed
- BsmtCond_3.0 : Basement condition TA Typical - slight dampness allowed
- LotShape_IR3 : General shape of property - Irregular
- Exterior2nd_Brk Cmn : Exterior covering on house - BrkComm Brick Common
- SaleCondition_AdjLand : Condition of sale - AdjLand Adjoining Land Purchase

Conclusion: all top 5 features are different from previous model. With original top 5 feature no more present now in the data, we can see that model picked up 'BsmtFinType2_4.0' which was originally at 8th position, but rest of the features are from much lower ranks

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 4

As per, Occam's Razor, in trying to understand something, getting unnecessary information out of the way is the fastest way to the truth or to the best explanation.

In statistical context, when you have two competing models that fit the data equally well, Occam's razor recommends to 'shave away all but what is necessary'. The concept of parsimony is based on Occam's razor, which also proposes that the model with fewer parameters to be preferred to the one with more.

So given two models that show similar 'performance' in the finite training or test data, we should pick the one that needs fewer parameters due to following reasons:-

- Simpler models are usually more 'generic' and are more widely applicable
- Simpler model require fewer training samples for effective training than the complex models and hence they are easier to train.
- Simpler models are more robust
    - Complex models tend to change wildly with changes in the training data set
    - Simple models have low variance, high bias and complex models have low bias, high variance
    - Simpler models make more errors in the training set, but can manage unseen data more efficiently
    - Complex models lead to overfitting - they work very well for the training samples however, they fail miserably when applied to other test samples and unseen data

Therefore, to make more robust and generalizable model, we need to make the model simple but not simpler that is not of any use in predicting the target variable. This is where regularization comes to help.

Regularization can be used to make the model simple. Regularization helps to strike the delicate balance between keeping the model simple and at the same time not making it too naive to be of any use.

For regression, regularization involves adding a regularization term to the cost which is minimized while learning the coefficients. For ridge model this term is the squares of the coefficients of the model while for Lasso this term is the absolute values of coefficients.

Also, Making a model simple leads to Bias-Variance Trade-off:
- A complex model will need to change for every little change in the dataset because it as practically memorized the training data instead of learning the patterns. Such model will be very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts 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 or unseen data. A complex model can do an accurate prediction provided there is enough training data. Models that are too naïve, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as its expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself with respect to changes in the training data.

Thus accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error as shown in the below graph.