Q 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 values obtained for both models are below:
   i. Lasso : 0.0001
   ii. Ridge : 50

b. Doubling the alpha values for Lasso regression
   i. RoofMatl_Membran no longer is significant and is replaced by RoofMatl_Tar&Grv among top 5 significant vars
   ii. Impact on R-squared Test

R-squared Score	α = 50	α = 100
Train Set	0.946773571	0.941567989
Test Set	0.838384567	0.845277679

c. Doubling the alpha values for Ridge regression:
   i. Variable significance altered for variables Neighborhood_NWAmes and GrLivArea from 2 and 3 to 3 and 2
   ii. Impact on R-squared test

R-squared Score	α = 50	α = 100
Train Set	0.869692823	0.855605987
Test Set	0.864978852	0.859138996

d. Most important predictor variables:

Model	Feature	Coefficient	Feature	Coefficient
Lambda	**α = 0.0001**		**α = 0.0002**	
Lasso	RoofMatl_WdShake	8.257163	RoofMatl_WdShake	6.468186
	RoofMatl_CompShg	7.950307	RoofMatl_CompShg	5.709007
	RoofStyle_Shed	7.488787	RoofStyle_Shed	5.673159
	RoofMatl_Membran	7.434091	RoofMatl_Tar&Grv	5.574498
	RoofMatl_Roll	7.394614	RoofMatl_Roll	5.471063
	RoofMatl_Tar&Grv	7.393598	RoofMatl_Metal	5.423408
	RoofMatl_Metal	7.320774	RoofMatl_Membran	5.328256
	Condition2_Norm	0.404744	Neighborhood_NWAmes	0.447003
	Neighborhood_NWAmes	0.399701	Condition2_Norm	0.32723
	2ndFlrSF	0.351742	MSZoning_RM	0.319149
Lambda	**α = 50**		**α = 100**	
Ridge	OverallQual	0.216431	OverallQual	0.213433
	Neighborhood_NWAmes	0.185092	GrLivArea	0.153806
	GrLivArea	0.172141	Neighborhood_NWAmes	0.118828
	Neighborhood_NoRidge	0.160588	2ndFlrSF	0.116329
	2ndFlrSF	0.139968	Neighborhood_NoRidge	0.110145
	BsmtCond_TA	0.138938	BsmtCond_TA	0.10386
	Condition1_Feedr	0.118399	GarageCars	0.096961
	GarageCars	0.110705	Condition1_Feedr	0.095007
	Neighborhood_CollgCr	0.102159	1stFlrSF	0.079674
	BsmtFinType1_BLQ	0.080653	BsmtFinType1_BLQ	0.074047

Q 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?

- R-squared on train and test sets for lasso are 0.946773571 & 0. 838384567
- R-squared on train and test sets for ridge are 0.869692823 & 0.864978852
- Given the fall in r2 score between train and test sets, ridge seems to be performing well as the r2 scores difference among train and test is very less (and this is even though r2 score on train is higher for lasso and test is higher with ridge)
- Considering above points on model accuracy, I'll go with ridge

Q 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?

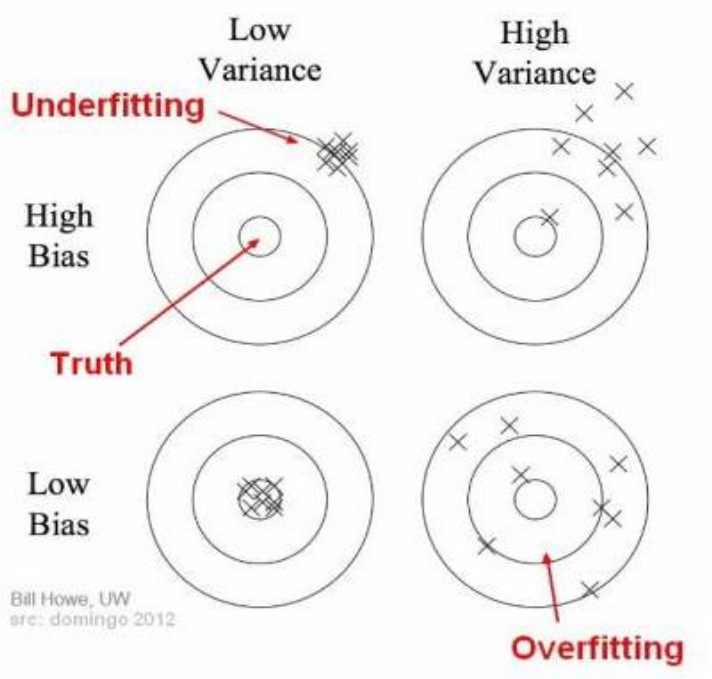
- The new top 5 significant features for lasso are below:

Feature	Coefficient
RoofMatl_CompShg	0.724111
Neighborhood_NWAmes	0.545512
Neighborhood_NoRidge	0.362281
GrLivArea	0.327994
Neighborhood_CollgCr	0.302222

Q 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?

- From Occam's Razor theory, a simpler model should be preferred over the more complicated models. Models should not be made more complex than necessary even if it means a hit on the accuracy of the model.
- Another theory to utilize while model building depends on the bias-variance trade-off concept. We need to minimize the total error while making sure we have good balance between bias and variance.

- The relation between model fit w.r.t bias and variance is as below:



  -