

## 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?**

Optimal value for alpha for lasso regression is coming as 0.001

Optimal value for alpha for ridge regression is coming as 0.01

If we double the values of alpha, model will try to penalize more and try to make more efficient coefficient values i.e. model will be more generalized and more simple.

Ridge - 0.01			Ridge - 0.02		
	Variable	Coeff			
0	MSSubClass	7.588	0	MSSubClass	8.020
109	RoofMatl_Metal	2.746	109	RoofMatl_Metal	2.304
110	RoofMatl_Roll	2.597	114	Exterior1st_AsphShn	2.190
111	RoofMatl_Tar&Grv	2.552	111	RoofMatl_Tar&Grv	2.186
114	Exterior1st_AsphShn	2.534	110	RoofMatl_Roll	2.170
112	RoofMatl_WdShake	2.489	108	RoofMatl_Membran	2.132
108	RoofMatl_Membran	2.479	112	RoofMatl_WdShake	2.103
113	RoofMatl_WdShngl	2.348	113	RoofMatl_WdShngl	1.994
9	BsmtFinSF2	0.852	9	BsmtFinSF2	0.711
3	OverallQual	0.643	3	OverallQual	0.606

  

Lasso -0.0001			Lasso -0.0002		
	Variable	Coeff		Variable	Coeff
0	constant	9.478	0	constant	10.547
15	GrLivArea	1.242	15	GrLivArea	1.187
114	RoofMatl_WdShngl	1.126	4	OverallQual	0.465
108	RoofMatl_CompShg	1.037	37	MSZoning_RL	0.261
111	RoofMatl_Roll	1.000	36	MSZoning_RH	0.246
112	RoofMatl_Tar&Grv	0.974	35	MSZoning_FV	0.240

113	RoofMatl_WdShake	0.949	5	OverallCond	0.238
109	RoofMatl_Membran	0.931	38	MSZoning_RM	0.222
110	RoofMatl_Metal	0.854	3	LotArea	0.205
4	OverallQual	0.426	25	GarageArea	0.199
37	MSZoning_RL	0.340	114	RoofMatl_WdShngl	0.197
36	MSZoning_RH	0.333	236	SaleType_ConLD	0.160
35	MSZoning_FV	0.329	16	BsmtFullBath	0.146
38	MSZoning_RM	0.300	58	Neighborhood_Crawfor	0.140
3	LotArea	0.282	108	RoofMatl_CompShg	0.115

### 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?**

I have chosen Lasso regression over ridge since Lasso helps in variable selection as it brings coefficient of variables to zero as well as shrinkage of coefficients where as ridge regression only tunes the parameters to reduce the value of coefficients.

With Lasso regression number of variables decreased which made model more simple and robust and helps in avoiding overfitting.

### 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?**

1. RoofMatl\_WdShake
2. RoofMatl\_Membran
3. RoofMatl\_Metal
4. OverallQual
5. MSZoning\_RL

### 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?**

We say model is robust and generalisable if test accuracy is not very different to train accuracy. Usually, simpler models are more robust though their accuracy will decrease in the process.

This can be understood from Bias-variance trade off graph. Simpler is the model , higher the Bias and lower the variance. Similarly, if model is more complex Bias will be less, but variance would be high.

In order to achieve generalisable model, too much weightage should not be given to outliers or noise. In order to do achieve robust and generalisable model we should be able to remove the independent variables that add more noise in the model rather than adding accuracy. If the model is not robust, it can't be trusted for predictive analysis.