

### 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?**

Answer:

- Alpha for:
  - Ridge: 1
  - Lasso: 0.001
- Ridge: Most of the scores including R2 remains similar/comparable even after doubling the alpha
- Lasso: Both R2 for train and test dropped by ~2% after doubling the alpha
- Order and coef values of important features changed in Lasso on doubling action while order remained same for Ridge
- Coef values changed for both Lasso and Ridge
- 'GrLivArea' remains the topmost feature even after changing both alphas

(Calculations present in jupyter notebook)

---

### 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:

- I'll choose Lasso regression because:
    - R2 values for train and test are very similar in Lasso (compared to Ridge which has considerable gap).
    - Lasso predicted features are something we figured out using Univariate analysis those were linearly aligned with target variable
-

### 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:

After re-running the Lasso model post removing the top 5 features from training and test sets,

We got following 5 features as new top:

1stFlrSF, 2ndFlrSF, Neighborhood\_NridgHt, Neighborhood\_Somerst, YearRemodAdd

(Calculations present in jupyter notebook)

---

### 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:

Model should be simple at the same time not as simple to underfit. Also, not so complex that it overfits.

We have to tradeoff between Variance and Bias.

Regularization can be done to make model more robust and generalisable. We have been taught 2 in course: Ridge and Lasso.

Regularization may reduce accuracy of training data a bit, however, accuracy of the model on test data will be significantly improved as – regularization makes sure that only important parameters/features are selected.