## Subjective Questions

**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?

**Ans.** The optimal values of alpha are,

Ridge = 0.4

Lasso = 0.0001

Top 5 features (** Taking only the positive coefficients)

| # | Ridge (Alpha = 0.4) | Lasso (Alpha = 0.0001) |
|---|---|---|
| 1 | GrLivArea | GrLivArea |
| 2 | 1stFlrSF | OverallQual |
| 3 | PoolQC_Ex | BsmtFinSF1 |
| 4 | BsmtFinSF1 | TotalBsmtSF |
| 5 | TotalBsmtSF | RoofMatl_WdShngl |

After running the models with double alpha, the top 5 features become

| # | Ridge (Alpha = 0.8) | Lasso (Alpha = 0.0002) |
|---|---|---|
| 1 | GrLivArea | GrLivArea |
| 2 | 1stFlrSF | OverallQual |
| 3 | 2ndFlrSF | RoofMatl_WdShngl |
| 4 | OverallQual | Neighborhood_NoRidge |
| 5 | RoofMatl_WdShngl | BsmtFinSF1 |

**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?

**Ans.** I will prefer to use Lasso regression as it reduces a number of features and makes the model simpler.

**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?

**Ans.** After rerunning the Lasso regression model with the original alpha value of 0.0001 and dropping the top 5 features the new top 5 features or predictor variables are,

1. 1stFlrSF
2. 2ndFlrSF
3. LotArea
4. PoolQC_Ex
5. Neighborhood_NoRidge

**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?

**Ans.** There are various ways to make a model robust and generalisable, e.g. handling outliers in data, use feature selection and regularisation methods like RFE, Ridge and Lasso, avoid overly complex models.

The implications of these is that the accuracy most likely dips by a very small percentage since some of the features are removed from the model. Hence the effect of these features on the model, however miniscule those are, are removed.