Question 1

The optimal value for lambda for Ridge (0.7) and Lasso (50) have both been calculated. If the value of each lambda is doubled then the performance and the results also change.

First looking at the performance

Metric	Ridge Regression	Ridge Regression Double Lambda	Lasso Regression	Lasso Regression Double Lambda
R2 Score (Train)	0.86	0.86	0.86	0.86
R2 Score (Test)	0.85	0.85	0.86	0.86
RSS (Train)	9.301155e+11	9.462081e+11	9.405387e+11	9.737830e+11
RSS (Test)	3.304262e+11	3.330065e+11	3.178295e+11	3.156826e+11
RMSE (Train)	28,501.37	28,746.87	28,660.62	29,162.74
RMSE (Test)	33,930.97	34,063.20	33,277.92	33,165.33

Most of the measures of error increase with the exception of the RSS and RMSE for lasso. The double lambda lasso model is performing better in some metrics. This is because we optimised using negative mean absolute error.

There were also some changes in the features identified as important. All the models identified the following top 5 features:

- GrLivArea
- OverallQualSq
- LotArea
- BsmtFinSF1
- Neighborhood_StoneBr

Apart from the Ridge model with the lambda doubled, this identified GarageCars in 5^{th} position instead of Neighborhood_StoneBr.

The other major difference is that the lasso model with the doubled value of lambda was simpler. It eliminated 5 additional features.

Question 2

Based on the answer to the first question I would probably apply the lasso model with the double lambda. It is a simpler model as it uses less features and it performs slightly better on some of the error metrics. As the model is simpler, it is more generalised, and I'd expect it to perform better on unseen data.

Question 3

If it is identified that the top 5 features are not available, I would have to recreate the model excluding these. To do this I would remove the features from an early stage. I would do this from before we perform feature elimination. This is because there may be other features that were correlated to the (no longer available) important features may have previously been eliminated. After this process the following top 5 features were identified:

- 1stFlrSF
- 2ndFlrSF
- GarageCars
- MasVnrArea
- ExterQualSq

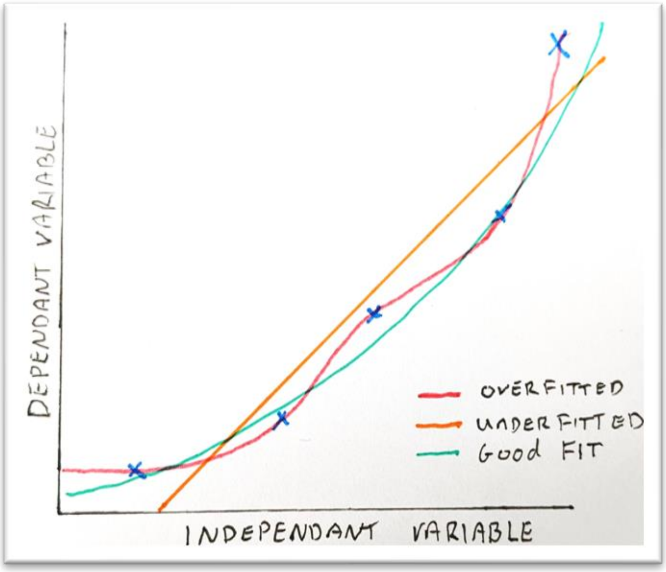
This makes sense as you would expect 1st Floor surface area plus 2nd Floor surface area to be very similar to above grade living area.

Question 4

If the model is robust the accuracy will be high during fitting as it builds a complex model that accounts for every feature in the training data. The downside is that it may not perform well on unseen data. This is because it may have overfitted, it has learned all the features but also all the noise in the data.

A generalised model will be very simple and based on a few features. If a model is too generalisable the accuracy of predictions will suffer. This is because the model is too simple (underfitted) and doesn't reflect all the features that impact the dependant variable.

To build a robust and generalisable model we need to balance the two attributes.



To achieve this we can use linear regression models that include regularisation. These are models that include a penalty in the cost function which is based on the model complexity. The penalty effects the way the optimal coefficients are identified . A lambda parameter is applied which is used to scale the penalty.

Hyper parameter tuning can then be used to find the optimal model. This involves training the model multiple times with different values of lambda and evaluating the performance each time. The optimal value of lambda will be the highest possible value that still achieves the highest observed accuracy.