Problem Statement - Part II

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?

Solution:

- The optimal value of alpha for ridge regression : 7.0
    - Model Evaluation for Ridge Regression with alpha=7.0
        - On Train dataset:
            - R2-score: 0.9165
            - RMSE: 0.1169
        - On Test dataset:
            - R2-score: 0.7870
            - RMSE: 0.1743

- The optimal value of alpha for lasso regression : 0.0001
    - Model Evaluation for lasso regression with alpha=0.0001
        - On Train dataset:
            - R2-score: 0.9163
            - RMSE: 0.1171
        - On Test dataset:
            - R2-score: 0.7877
            - RMSE: 0.1740

If we choose double the value of alpha for both ridge and lasso regression:
For ridge, it will now try to apply more penalties to the model and thus would try to make the model simpler and more generalised.

- New optimal value of alpha for ridge regression : 14.0
    - Model Evaluation for Ridge Regression with alpha=14.0
        - On Train dataset:
            - R2-score: 0.9164
            - RMSE: 0.1170
        - On Test dataset:
            - R2-score: 0.7891
            - RMSE: 0.1734
    - There is no significance difference on R2-score but we can observe slightly lower R2-score on train data but a slightly higher R2-score on test data
    - There is no significance difference for RMSE as well but we can observe it is slightly lower on test data

For lasso, it will try to penalize the model and as a result, more model coefficient reduces to zero. As we increase the alpha r2-score will going to decrease.

- New optimal value of alpha for lasso regression : 0.0002
    - Model Evaluation for Lasso Regression with alpha=0.0002

        - On Train dataset:
            - R2-score: 0.9166
            - RMSE: 0.1168
        - On Test dataset:
            - R2-score: 0.7842
            - RMSE: 0.1754
    - There is no significance difference on R2-score but we can observe slightly lower R2-score on test data
    - There is no significance difference for RMSE as well but we can observe it is slightly higher on test data

- The most important feature for ridge with alpha = 14.0 are:
    - OverallQual
    - 1stFlrSF
    - 2ndFlrSF
    - BsmtFinSF1
    - GrLivArea
- The most important feature for lasso with alpha = 0.0002 are:
    - 1stFlrSF
    - OverallQual
    - 2ndFlrSF
    - BsmtFinSF1
    - OverallCond

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?

Solution:

- The optimal value of alpha for ridge regression : 7.0
- The optimal value of alpha for lasso regression : 0.0001

As discussed above in detail, we can observe there is no significance difference between the two models but lasso regression is showing slightly higher R2-score on test data than ridge regression.

I would choose lasso regression as it helps in feature selection (by pushing coefficient equals to zero) which results in model performance. Lasso regression also penalizes the model on new feature addition.

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?

Solution:

- The final model’s 5 most important predictor variables in the lasso with alpha = 0.0001:
    - 1stFlrSF
    - OverallQual
    - 2ndFlrSF
    - OverallCond
    - BsmtFinSF1

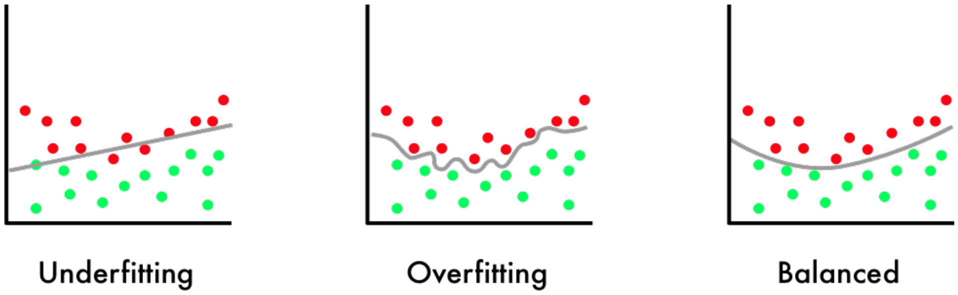
After building a new model on removing original 5 important predictors we got below new 5 most important predictor variables in the lasso:

- GrLivArea
- ExterQual
- GarageArea
- CentralAir
- BsmtQual

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?

Solution:

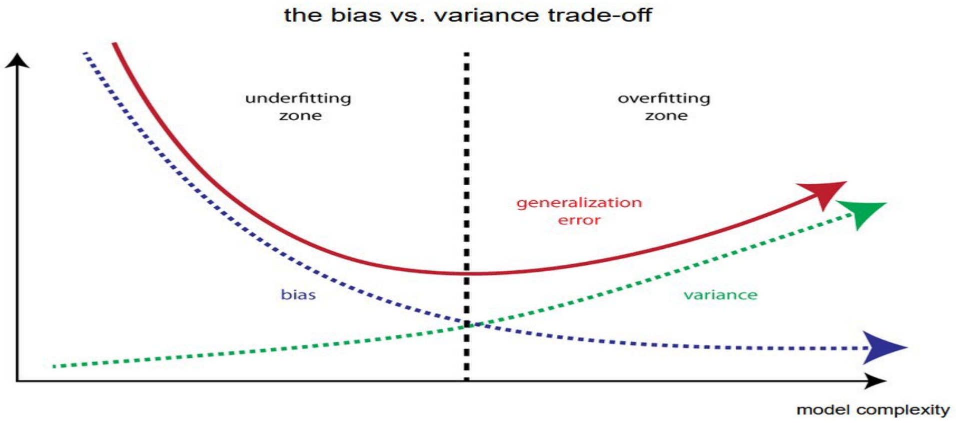
- To make a model robust and generalisable, we need to cater a trade-off between bias and variance. Theoretically, we generally target a model with low bias and low variance but practically it is hard to achieve that.



- Let’s look at the aspect if we do not follow a trade-off between bias and variance
    - High variance (Overfitting):
        - High variance models puts lots of efforts and attention to the training data as a result model is not generalised enough to accommodate or perform better on unseen data
        - Therefore such models perform very well on training data and result in high error rate on unseen/test data.
        - This is a clear case of overfitting.
    - High bias (Underfitting):
        - High bias models puts very little attention on training data as a result model is over-simplified.
        - Therefore such models results in high errors on both training as well as test data.

        - This is a clear case of underfitting.
    - Low bias and low variance (Balanced):
        - For a balanced model we need to make simple yet robust and generalised model
        - We need to find a balance between bias and variance (trade-off), so that we can minimises the total error.

- To achieve low bias and low variance in a model we apply regularisation on that. Ridge and lasso regression are few of the techniques to minimises the model complexity and avoids high variance(overfitting) and high bias(underfitting).



- Ridge regression:
    - Here in ridge regression models cost function altered or controlled by adding penalty which is equal to squares of the magnitude of coefficient.
    - The ridge regression shrinks the coefficient by applying a penalty term i.e. lambda.
    - This shrinkage helps to minimise the model complexity and multicollinearity.
    - When compare to lasso, model performance could be poor as it continues with all the features
- Lasso regression:
    - Here in lasso regression models cost function altered or controlled by adding penalty which is equal to absolute of the magnitude of coefficient.
    - The lasso regression not only shrinks the coefficient by applying penalty term(lambda) but also does a feature selection by reducing coefficient equals to zero.
    - Model performance could be good as it does a features selection.

- We should understand the bias and variances in the model at hand so that we could get acceptable accurate results(or minimize the error terms) by allowing necessary compromises(trade-off) while building a model.