

Answer 1: The optimal value of alpha for ridge and lasso regression are: 7.0 and 0.0001 respectively. As the value of lambda doubles for ridge and lasso regression, the variance decreases and the bias increases. The most important predictor variables after doubling the value of lambda are:

1. For Ridge Regression:

OverallQual_9, Neighborhood_StoneBr, GrLivArea, FullBath_3, BsmtExposure_Gd.

2. For Lasso Regression:

GrLivArea, OverallQual_10, OverallQual_9, TotalBsmtSF, Neighborhood_StoneBr.

Answer 2: I will choose to apply Lasso Regression because it helps in feature selection by eliminating most of the features by making their coefficients as 0 and also it has a r^2 _score of 0.88 on test set.

Answer 3: After creating another model excluding the five most predictor variables, the new five most important predictor variables now are:

1stFloorSF, BsmtFinSF1, 2ndFlrSF, LotArea, TotRmsAbvGrd_12.

Answer 4: Building a model such that the model has least effect from the outliers ensures that the model is robust and generalisable. As the model becomes more generalised, its accuracy stays intact when it is tested for the unseen data because the model will not overfit.