

Assignment-based Subjective Questions

- 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?

Observation:

- The optimal value of alpha for ridge – 1.0 and lasso – 20
- In the case of ridge regression,
 - After doubling the alpha value in ridge, the coefficients reduced after doubling, r2 scores reduced and rss increased.
- In the case of lasso regression,
 - After changing the alpha values, the coefficients decreased after doubling, r2 scores reduced and rss and mse decreased
- In case of lasso, the predictor variables changes when alpha changes
 - TotalBsmtSF
 - OverallQual
 - 2ndFlrSF
 - Exterior1st_BrkComm
 - Exterior1st_AsphShn
 - YearBuilt
- In case of ridge, the predictor variables are,
 - OverallQual
 - MasVnrArea
 - TotalBsmtSF
 - GarageCars
 - LotFrontage

- 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

Observation: I would choose to apply Lasso regression with optimal value of 20, as the R² score is high as (88.22%) with less RSS score compared to when i choose alpha 1 or 10

- 3) After building the model, you realized 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?

Observation:

After removing the previous predictor variables from the data set and try to apply the lasso on the remaining one, i saw the below predictor variables now

- FullBath
- Neighborhood_MeadowV
- BsmtUnfSF
- Neighborhood_Mitchel
- BsmtFinSF1

- 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?

- a) A model is considered to be robust if its output dependent variable is consistently accurate even if one or more of the input independent features or assumptions are drastically changed due to unforeseen circumstances.
- b) When the r² score are consistent during the training of the model and the testing, Also depends on the RFE values and RSS values
- c) The accuracy of the model varies with the type and number of feature set input to it. A change in the model input feature set can change the different metrics that are used as a measure of efficiency. Adjusted R Squared metric can be used to explain how well independent features explain the variability in the dependent variable/ target variable. For

regression models, there are many metrics that provide information on the model's performance. A robust model will show lower susceptibility to changes in the feature sets.