

Subjective Questions Solutions

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?

Answer:-

- Optimal Value of Ridge Regression is 7.
- Optimal Value of Lasso Regression is 0.0005.

Ridge

'LotArea': 0.0226,
'OverallQual': 0.071,
'OverallCond': 0.0438,
'BsmtFinSF1': 0.028,
'TotalBsmtSF': 0.0501,
'1stFlrSF': 0.0154,
'2ndFlrSF': 0.0145,
'GrLivArea': 0.0812,
'BsmtFullBath': 0.0116,
'FullBath': 0.0129,
'HalfBath': 0.0228,
'Fireplaces': 0.0213,
'GarageCars': 0.035,
'WoodDeckSF': 0.0099,
'isRemodel': -0.0104,
'BuiltOrRemodelTime': -0.0194,
'NewOrOldGarage': 0.0155,
'd_BsmtExposure': 0.0182,
'd_BsmtFinType1': 0.0083,
'd_HeatingQC': 0.0198,
'd_KitchenQual': 0.0172,
'd_BldgType': -0.0197,
'd_SaleCondition': 0.0147,

'MSZoning_FV': 0.0618,
'MSZoning_RH': 0.0242,
'MSZoning_RL': 0.0913,
'MSZoning_RM': 0.0654,
'Neighborhood_Crawfor': 0.0259,
'Neighborhood_Edwards': -0.0103,
'Neighborhood_MeadowV': -0.0083,
'Neighborhood_NridgHt': 0.0239,
'Neighborhood_OldTown': -0.0155,
'Neighborhood_SWISU': -0.0083,
'Neighborhood_StoneBr': 0.0178,
'Exterior1st_BrkComm': -0.0096,
'Exterior1st_CemntBd': -0.0123,
'Exterior1st_Stucco': 0.0135,
'Exterior1st_VinylSd': -0.0395,
'Exterior1st_Wd Sdng': -0.0311,
'Exterior2nd_CmentBd': 0.013,
'Exterior2nd_Stucco': -0.0185,
'Exterior2nd_VinylSd': 0.0354,
'Exterior2nd_Wd Sdng': 0.0201,
'Foundation_CBlock': 0.0195,
'Foundation_PConc': 0.044,
'Foundation_Slab': 0.0137,
'Foundation_Stone': 0.0063,
'GarageType_CarPort': -0.0109,
'GarageType_Detchd': -0.009,
'GarageType_No Garage': -0.0114

Lasso:-

'LotArea': 0.0221,
'OverallQual': 0.0732,
'OverallCond': 0.0437,
'BsmtFinSF1': 0.029,
'TotalBsmtSF': 0.0504,
'1stFlrSF': 0.0031,
'2ndFlrSF': -0.0,
'GrLivArea': 0.0981,
'BsmtFullBath': 0.0107,
'FullBath': 0.0112,

'HalfBath': 0.0216,
'Fireplaces': 0.0212,
'GarageCars': 0.0358,
'WoodDeckSF': 0.0097,
'isRemodel': -0.0102,
'BuiltOrRemodelTime': -0.0196,
'NewOrOldGarage': 0.0147,
'd_BsmtExposure': 0.018,
'd_BsmtFinType1': 0.0074,
'd_HeatingQC': 0.0192,
'd_KitchenQual': 0.0169,
'd_BldgType': -0.0189,
'd_SaleCondition': 0.0146,
'MSZoning_FV': 0.0627,
'MSZoning_RH': 0.024,
'MSZoning_RL': 0.0938,
'MSZoning_RM': 0.0657,
'Neighborhood_Crawfor': 0.0252,
'Neighborhood_Edwards': -0.0097,
'Neighborhood_MeadowV': -0.0074,
'Neighborhood_NridgHt': 0.0226,
'Neighborhood_OldTown': -0.0152,
'Neighborhood_SWISU': -0.0086,
'Neighborhood_StoneBr': 0.0168,
'Exterior1st_BrkComm': -0.0093,
'Exterior1st_CemntBd': -0.0,
'Exterior1st_Stucco': 0.0123,
'Exterior1st_VinylSd': -0.0278,
'Exterior1st_Wd Sdng': -0.0285,
'Exterior2nd_CmentBd': 0.0008,
'Exterior2nd_Stucco': -0.0172,
'Exterior2nd_VinylSd': 0.0241,
'Exterior2nd_Wd Sdng': 0.0168,
'Foundation_CBlock': 0.017,
'Foundation_PConc': 0.0421,
'Foundation_Slab': 0.0126,
'Foundation_Stone': 0.0057,
'GarageType_CarPort': -0.0102,
'GarageType_Detchd': -0.0085,
'GarageType_No Garage': -0.0109

Variables with highest correlation is good predictor. Compare all coefficients to select best predictor. We can also look into change in R-square value.

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?

Answer:-

Among Ridge and Lasso Regression, Lasso is the best option to choose because in regression, Lasso Regression helps in best columns reduction and also robust in nature.

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?

Answer:-

The five most important predictor variables/attributes are:-

- Zoning Classification
- Total Basement in Square Feet
- Foundation Type of House
- Overall Quality and Condition of the house
- Basement Finished Square Feet

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?

Answer:-

If we have to choose between two similar performance based model in finite train and test data. Using Occam's Razor, we should choose the model that makes lesser error on test data because simple model are more generic and are applicable widely. These models require less training samples for training. Thus making it easier to train. Simple models are robust because unlike them, complex models will widely fail if we make changes in training data set. Simpler models make more errors in the training set. Complex Model makes less error in the training set but will fail miserably when we apply other test sample which is also known as Over fitting.

Therefore to make model robust and generalizable, make the model simple but not too simpler which will be not used at all.

The model should be accurate for the datasets other than that which were used in training. Too much importance should not be given to the outliers so that the accuracy predicted by the model is too high. Therefore, the outlier analysis needs to be done and only those which are relevant to the dataset need to be kept.