Subjective Questions and Answers (Problem Statement)

by Himanshu S (2024)

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

The optimal alpha for Lasso Regression is approximately 0.00077, and for Ridge, it is 17.886. If we were to choose double the value of alpha for both Ridge and Lasso, it would lead to stronger regularization, resulting in a model with fewer features. The coefficients of th e re maining fe atures wo uld li kely be further shrunk towards zero. After the change, the most important predictor variables would be those with non-zero coefficients. Af ter im plementing th e ch anges in Lasso Regression, the critical predictor variables are:

1. GrLivArea
2. OverallQual

After implementing the changes in Ridge Regression, the key predictor variables are:

1. MSZoning_FV
2. MSZoning_RL
3. Neighborhood_Crawfor
4. MSZoning_RH
5. MSZoning_RM
6. SaleCondition_Partial

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

The R-squared values for Lasso Regression (0.916) and Ridge Regression (0.918) based on the R2 score show that the Ridge model is slightly better, but this is subjective. If we have many features that are highly correlated and want to keep all of them in the model, we go for Ridge regression. It handles multicollinearity by shrinking correlated features towards each other without excluding any of them. If we suspect that some features are not really useful or want a simpler model with fewer features, then Lasso regression is the better choice. Lasso tends to set some coefficients exactly to zero, which means it performs feature selection by excluding certain features.

Question 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?

Answer

We find that the next most important variables are: For Ridge:

1. OverallCond: Overall condition of the house
2. BsmtFullBath: Basement full bathrooms
3. BsmtFinSF2: Type 2 finished square feet
4. 2ndFlrSF: Second floor square feet
5. 1stFlrSF: First Floor square feet

For Lasso:

1. BsmtFullBath: Basement full bathrooms
2. OverallCond: Overall condition of the house
3. BsmtFinSF2: Type 2 finished square feet
4. BsmtFinSF1: Type 1 finished square feet
5. OverallQual: Rates the overall material and finish of the house

Question 4

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer

We should choose a model that balances simplicity and accuracy to ensure robustness and generalizability. When making a model, the bias-variance trade-off is crucial, where a simpler model with higher bias and lower variance tends to be more generalizable. Striking this balance prevents overfitting (memorizing the entire dataset) and underfitting (not being able to understand the data patterns at all), allowing the model to perform well on both training and test datasets. Managing bias and variance ensures reliable predictions of our model across diverse datasets, test, and production data, emphasizing the importance of finding the right equilibrium for building effective machine learning models.