Subjective Questions Answers

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:

<table>
<tr>
<td>What is the optimal value of alpha for ridge and lasso regression?</td>
<td>Optimal values for Lasso and ridge are:
1. Ridge: 0.4
2. Lasso: 0.001</td>
</tr>
<tr>
<td>What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?</td>
<td><u>Ridge:</u>
Observed that the beta coefficients for the model will be incremented. While some variables betas doubled some betas change moderately. Also, the top variables for model changed.
<u>Lasso:</u>
Observed that, the beta coefficients for the variables will be incremented. While some variables betas increased by 0.1, some betas decreased drastically low value after doubling. Also, the top variables for model will changed.</td>
</tr>
<tr>
<td>What will be the most important predictor variables after the change is implemented?</td>
<td><u>Ridge top 10 variables: Before Doubling:</u>
<table>
<tr><td>Features</td><td>Coef</td></tr>
<tr><td>PoolQC_Ex</td><td>0.871426</td></tr>
<tr><td>PoolQC_Fa</td><td>0.646504</td></tr>
<tr><td>Condition2_Norm</td><td>0.524929</td></tr>
</table>
<u>Ridge Top-10 variables: After Doubling:</u>
<table>
<tr><td>Features</td><td>Coef</td></tr>
<tr><td>PoolQC_Ex</td><td>0.705534</td></tr>
<tr><td>OverallQual_9</td><td>0.46611</td></tr>
<tr><td>PoolQC_Fa</td><td>0.437643</td></tr>
</table>
<u>Lasso Top-10 variables: Before Doubling:</u></td>
</tr>
</table>

Subjective Questions Answers

Features	*Coef*
OverallQual_9	*0.452963*
GrLivArea	*0.343489*
Neighborhood_Crawfor	*0.297178*

Lasso Top-10 variables: After Doubling:

Features	*Coef*
OverallQual_9	*0.388706*
GrLivArea	*0.335896*
Neighborhood_Crawfor	*0.280437*

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:

which one will you choose:Ridge/Lasso?	Regularization method selected is Lasso because of low Negative MSE when compared with Ridge regularization. More importantly, Lasso regression shrinks the coefficients to zero which makes variable selection easy which is equivalent to the particular feature being excluded from the model and it helps to reduce the model complexity and multi-collinearity.

Subjective Questions Answers

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:

Below are the top 5 variables after removing the top 5 variables from the first model and remodeling the dataset using Ridge and Lasso regularization.

We chose lasso again as the MSE for lasso is low when compared with Ridge resulting in better variable selection as it shrinks the variables to zero resulting in variable selection..

Top 5 Variables:

1. 2ndFlrSF
2. 1stFlrSF
3. OverallCond_9
4. KitchenQual_Ex
5. Neighborhood_Somerst

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:

How can you make sure that a model is robust and generalisable	Robustness for a model means it should be resilient to noise of unseen data. When the model testing error is least and does not deviate too much from the training error it is a robust model. Generalization means, even though we get some noise in the unseen dataset the model performs in a stable manner is considered generalizable model. Moreover, a model, which is '**Simple'**, is usually more generic

Subjective Questions Answers

	and generalizes more efficiently, that the model which is complex; a complex model, though it performs well on train data it fairs poorly on the unseen data, i.e. high test error. This means generalizability of model is poor. Using regularization techniques like ridge and lasso, we can ensure the robustness and genralizability of the modelas these techniques when applied performs shrinking or regularizing the coefficients, prediction accuracy can be improved, variance can be decreased, and model interpretably can also be improved. Moreover, if the model is capable of training well on the train dataset it will learn the noise associated with the dataset and could generalize the unseen data effectively. For instance, cross-validation is one such technique where the dataset is trained and tested on multiple chunks of the train dataset with building multiple models so that the training model learns well on the data and generalize well w.r.t unseen data. ***A model to be robust and generalizable should be simple and not simpler (generalization) and consistent with error terms on test and train data (robustness)***.
What are the implications of the same for the accuracy of the model and why?	Accuracy of the model will be highly improved when the model is stable and generalizing the unseen data. When the model is **not** simpler and not too complex, that means we have a extremely good trade-off between simple model and complex model. That means accuracy of the model will be significantly good.

.