

Part II

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?

Optimal alpha (lambda) for:

ridge regression: 500

lasso regression: 900

After doubling the alpha, r^2 score of unseen data for:

ridge regression fell from 87.7% to 86.7%

lasso regression fell from 88.5% to 87.8%

Most important predictor for:

ridge regression: OverallQual

lasso regression: GrLivArea

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?

I choose Lasso regression model for the following reasons:

- explains unseen data better
- less complex model

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?

GrLivArea: Above grade (ground) living area square feet

OverallQual: Rates the overall material and finish of the house

Neighborhood: Physical locations within Ames city limits

- NoRidge: Northridge

- NridgHt: Northridge Heights

RoofMatl: Roof material

- WdShngl: Wood Shingles

KitchenQual: Kitchen quality

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

We can make the model more generalized by increasing the hyper-parameter (λ/α) value. As λ increases, the penalty on the number of features increases. So the number of features that describe the data reduces. This reduces the accuracy of the model on the training data. However, the overall accuracy (training data + unseen data) becomes more consistent, i.e., the accuracy on unseen data will be more or less the same as the accuracy on training data.

As λ further increases, the overall accuracy decreases and the model fails to learn any patterns in the data.