

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?

Ans: Please find the answer in the notebook attached:

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?

Ans: Lasso regression would be a better option it would help in feature elimination and the model will be more robust.

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?

Ans: Please find the answer in the notebook attached.

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?

Ans: A model is considered to be robust if its output dependent variable (label) is consistently accurate even if one or more of the input independent variables (features) or assumptions are drastically changed due to unforeseen circumstances. A simple model is robust and does not change significantly if the training data points undergo small changes. A simple model may make more mistakes in the training set, but it is bound to outperform complex models when it sees new data. But we do not want to create a model too simple as it will not perform very well causing underfitting, so we would like to maintain the balance between the complexity and simplicity of the model to make it more generalisable and robust.

When the model is too complex, it needs to memorize a lot and therefore changes when introduced it is more likely to swing wildly even with small changes in training data. Accuracy of such model on test data will be deviated and will highly likely not be predicted correctly, i.e., they will have high variance.

The model that is too simple and has a very few parameters in it maybe high bias and low variance. If the model is not robust, it cannot be trusted for predictive analysis.