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:

In our analysis, the optimal regularization param was identified as 10 for ridge and 0.001 for lasso reg. if the alpha value for ridge is doubled, the model becomes more strict, pushing smaller Coeff close to zero. This can slightly hurt accuracy if some imp features lose their impact.

For lasso, increasing alpha to 0.002 strengthens its feature select behaviour, meaning more var may be dropped entirely. While this simplifies the model, it can also remove important predictors and reduce performance.

After changing alpha, the models would need to be re-trained to identify new key predictors , although the strongest features from the original models are likely to remain relevant.

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 decision between ridge and lasso depends on what we want to achieve. In our case, lasso is better choice because its performance is very close to ridge ,while also offering extra advantage of feature selection. By shrinking some Coeff all the way to zero , lasso reduces number of variables in the model , making it simpler, easy to interpret and less likely to over fit. It makes it a more practical option for this problem.

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:

if the original top five predictors from the lasso are missing in new data , the model must be rebuilt without them. To identify the most imp variables , we would retrain the lasso regression on uploaded dataset and analyse the resulting coeff. Since this is a hypothetical situation and the model has not actually been rerun, the exact new top predictors cannot be determined without performing this step.

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:

to make sure model is robust and generalisable , it should perform well both on training and unseen data. This can be achieved using techniques such as cross-validation , proper train-test splits, regularizations, feature selections and handling outliers. These methods prevent the model from learning noise and help it capture true patterns.

In terms of accuracy, a good model shows similar performance on new data as on training data. If it performs well only on training data, it is overfitting. If it performs , poorly on both , it is underfitting. The goal is to balance bias and variance to achieve reliable predictions on unseen data.