Advanced Regression Subjective Question by Abhishek Singh

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 1- The optimal value of alpha for ridge and lasso regression depend on the dataset and problem. Alpha in Ridge controls the strength of regularization, a large alpha push for a coefficient shrinkage. In Lasso, it determines feature selection, with higher alpha leading to fewer features. To find the most important predictors, fit the model with the new alpha and rank predictors based on the magnitude of non-zero coefficients. Cross-validation is crucial for selecting the right alpha for your data.

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 2 - I would choose Lasso regression if we have many features, some of which might not be very important. Lasso can help us select the most important features and simplify the model.I would choose Ridge regression if we're concerned about features being too related to each other (multicollinearity) and want to keep all features while reducing their impact.

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 3 – If the five most important predictor variables from your original Lasso model are not available in the incoming data, you cannot directly identify the new top five important variables without retraining the model on the updated data. To find the new most important predictor variables:

1. Exclude the original top five variables from your dataset.
2. Retrain the Lasso model on the remaining variables using the current data.
3. Examine the coefficients of the new model to identify the new set of important predictor variables based on the available data.

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 4 - To make sure a model works well in different situations, you should:

1. Train it on various examples.
2. Check its performance on different test sets.
3. Choose important factors and avoid irrelevant ones.
4. Use techniques to prevent extreme values.
5. Fine-tune its settings.
6. Test it on new data it hasn't seen before.

A good model might not be perfect on the training data, but it will do better in real-life situations because it's not too specialized.