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

In Ridge and Lasso regression, the parameter alpha(α) is the penalty term that denotes the amount of shrinkage that will be implemented in the equation. A larger value of alpha(α) penalize the optimization of function. If α is close to zero(0), the Ridge term itself is very small and the final error is based on RSS alone. If α is too large, the impact of shriking grow and the coefficient β_1, β_2 --- β_n tends to zero. The value of α is often chosen via cross validation by checking a bunch of different values on training data and seeing which yields the best R2 on test data.

If we double the value of alpha for both Ridge and Lasso regression, it will increase the penalty term. This means that the coefficient of your model will shrink more potentially towards zero. This shrinkage can lead to some coefficient becoming exactly zero in Lasso regression which is a form of automatic feature selection.

The most important predictor variable after this change would be those with non-zero coefficient. In other words, these are the variables that the model has found to be most predective even after increase the penalty term. This can vary depending on specific dataset and model.

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

The choice between Ridge and Lasso regression depends on the specific char--acteristics of our dataset.

Ridge regression is a good choice when we have a lot of feature an believe that all of them should have some effect on the output, even if it's small. Ridge regression will shrink ~~shrink~~ the coefficient towards zero, but it won't make them exactly zero. This means that all feature will still be in the model, but those with smaller effect will have smaller coefficients

Lasso regression, on the other hand, is useful when you believe that only a subset of your features are actually influencing the output. Lass has the ability to force some coefficients to be exactly zero, effectively performing feature selection's effects.

So If we want a model that includes all features but with reduced compexity, go for Ridge. If we want a sparser solution with fewer feature, go for Lasso.

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 we exclude the five most important predicter variable from the lasso model, the next set of important predictor will move up in terms of importance. Lasso tends to prioritize certain variables by shriking others to zero. After removing the top five, the next five in importance will take their place. The order and specific variables can be determined by retraining the model without initially identified top value.

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 ensure that a model is robust and generalizable, we can follow few steps:-

Use a diverse dataset !- The dataset should contain a wide range of examples from the problem space. This helps the model to learn the underlying patterns in the data and generalize well to unseen data.

Implement cross-validation !- Impliment cross-validain, During training, we can use cross validation technique like K-fold cross validation. This help in understanding how the machine learning model would generalize to an independent dataset.

* Apply Regularization technique !- Regularization technique like L1 & L2 regularization can help prevent over-fitting by adding a penalty term to the loss function that the model optimizes

Use early stopping:- Early stopping is a form of regularization used to avoid overfitting when training a learner with an iterative method, such as gradient descent. This is done by stopping training when the performance on a validation dataset starts to degrade.

Data augmentation:- Data augmentation techniques such as flipping, rotation, zooming etc., can be used to increase the amount of training data and help improve the model's ability to generalize