Advanced Regression subjective Questions

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 :

The optimal value of alpha for Ridge and Lasso regression depends on specific data and the problem you are trying to solve. Alpha is used and helps to determine the complexity of the model. There is always a trade off Between variance and bias and write alpha values manages both and is uses to have a optimum value of both , based on different alpha values the regression is run against. As in the case study the optimum value for lasso was chosen as 0.01 for lasso and 2 for ridge.

In Ridge regression, increasing alpha makes the model more regularized. It means that the model becomes more constrained, and the impact of each predictor on the outcome is reduced. If you double the value of alpha in Ridge regression, the model becomes even more constrained. The coefficients of the predictors will be pushed closer to zero, making the model simpler.
In Lasso regression, increasing alpha also increases the regularization effect. Lasso regression can even make some coefficients exactly zero, effectively performing variable selection. If you double the value of alpha in Lasso regression, more coefficients may be set to zero, resulting in a simpler model with fewer important predictors.

In Ridge regression, the coefficients will be smaller but still non-zero, while in Lasso regression, some coefficients may be exactly zero. So, in Lasso regression, the most important predictor variables are the ones that still have non-zero coefficients after the change.

Remember, choosing the right value of alpha depends on your data and problem. It's good practice to use techniques like cross-validation to find the best alpha value for your specific situation.

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:

regularize coefficients and improve the prediction accuracy also with the decrease in variance, and making the model simple and robust.

Ridge regression, uses a tuning parameter called lambda as the penalty is square of magnitude of coefficients which is identified by cross validation. Residual sum or squares should be small by using the penalty. The penalty is lambda times sum of squares of the coefficients, hence the coefficients that have greater values gets penalized. Wit increase in value of lambda the variance in model is dropped and bias remains constant. Ridge regression includes all variables in final model unlike Lasso Regression. Lasso regression, uses a tuning parameter called lambda as the penalty is absolute value of magnitude of coefficients which is identified by cross validation. With lambda value increases Lasso shrinks the coefficient towards zero and it make the variables exactly equal to 0. Lasso also does variable selection. When lambda value is small it performs simple linear regression and as lambda value increases, shrinkage takes place and variables with 0 value are neglected by the model.

So based on lasso and ridge performance , its advised to choose the model which is simpler .For example for our case study , we used lasso rather than ridge , though ridge performance was comparatively high , but lasso helped us choose a simpler model.

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 :

Important Predictor variables now are :

1. **BsmtFullBath** - Basement full bathrooms
2. **TotalBsmtSF** - Total square feet of basement area
3. **LotArea** - Lot size in square feet.
4. **PoolArea** – Pool Area in square feet.
5. **MSSubClass** - Identifies the type of dwelling involved in the sale.

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 ensure that a model is robust and generalizable, you can follow several practices:

1. **Train the model on diverse and representative data**
2. **Use cross-validation**
3. **Feature engineering and selection**
4. **Regularization techniques like lasso or ridge.**
5. **Hyperparameter tuning**

A highly accurate model may perform well on the training data, it may not generalize well to unseen data if it lacks robustness. Therefore, model accuracy alone is not a sufficient measure of its effectiveness. It's crucial to take care of robustness and generalization to ensure reliable performance on new, unseen data.