

Advanced Regression Assignment 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 to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Answer

- Optimal value of alpha for Ridge = 2.0 and Lasso = 0.0001
- After alpha value doubled i.e Ridge = 4 and Lasso = 0.0002, observed
 - For Ridge, R2 score decreased by 0.4% and increase in RMSE by 0.1%
 - For Lasso, R2 score decreased by ~0.7% and increase in RMSE by ~0.3%. It was able to eliminate 42 features with almost similar accuracy as earlier.
- Top 5 most important predictor variables after the alpha is doubled

Ridge	Lasso
GrLivArea	GrLivArea
1stFlrSF	OverallQual_9
BsmtFinSF1	BsmtFinSF1
OverallQual_9	Neighborhood_StoneBr
Neighborhood_StoneBr	OverallQual_8

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

- Optimal values of alpha for Ridge = 2.0 and Lasso = 0.0001
- Lasso regression model should be preferred as R2 score is better and RMSE is lower on test set. Additionally, it also helped in reducing the features by making 142 coefficients “0”. So, the model is simpler compared to Ridge or Base Linear Regression model and can be easily generalizable and 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?

Answer

- After dropping 'GrLivArea', 'OverallQual_9', 'BsmtFinSF1', 'Neighborhood_StoneBr', 'OverallQual_8', another model was built.
- Top 5 most import predictor variables with the new model are
 - 1stFlrSF
 - 2ndFlrSF
 - Neighborhood_Crawfor
 - LotArea
 - GarageArea

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

- As per **Occam's Razor**, having a simpler model will gives advantages like generalisability, robustness, fewer assumptions, and less data required for learning.
- As complexity of the model increases, accuracy increases, variance increases and bias decreases which leads to overfitting of the model. The model will not do well on unseen data.
- If the model is too simple, variance would be too low, accuracy will be low, but the bias will be very high, and the model would be underfitting
- Regularization of a model can help strike balance between keeping the model simple but not too naïve which in turn helps finding a balance between variance and bias bringing the total error down and having decent accuracy.