

Advance Regression – Subjective Questions

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?

Ans:- Optimal alpha value for ridge = 1.0

Optimal alpha value for losso = 0.0001

Following are observation changes in the model

- a. The accuracy of the model reduced

	Ridge		Ridge	
	optimal	double	optimal	double
Regression train	0.9	0.89	0.93	0.91
Regression test	0.87	0.87	0.86	0.86

- b. There is change in the order of predictive parameters used
- c. Observed little difference in the model co-efficients which became shrink

Predictor variables after changes – LotFrontage, OverallCond, GrLivArea, 1stFlrSF, MSSubClass, FullBath, 2ndFlrSF, YearRemodAdd, YearBuilt, MSZoning_RM

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?

Ans:- It is important to regularize coefficients and improve the prediction accuracy also with decrease in variance, and making the model interpretable. Below are the values obtained on training and testing data from ridge and lasso model. I will choose Ridge over Lasso due to the performance of testing is slightly high and also the accuracy between train and testing is acceptable.

	Ridge	Ridge
	optimal	optimal
Regression train	0.9	0.93
Regression test	0.87	0.86

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?

Ans:- The 5 most important predictor variables are

LotArea
ExterCond
Exterior1st_AsphShn
RoofMatl_Roll
RoofMatl_Metal

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?

Ans:- The model should be as simple as possible, though its accuracy will decrease but it will be more robust and generalisable. It can be also understood using the bias-variance trade-off. By applying regularization technique we can prevent overfitting and improve the model's generalization performance. Regularization can slightly reduce the accuracy of the model on the training data but often improves its accuracy on new, unseen data by reducing overfitting and improving generalisation.

Evaluating the model on unseen test data provides an unbiased estimate of its accuracy on new data. It helps ensure the model perform well in real-world scenarios.

—