

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 for Ridge is 1 and Lasso is 0.0001

If we double, then the alpha for Ridge is 2 and for Lasso is 0.0002

After the changes are implemented, since the alpha value is quite small, doubling it did not do any significant change in both the models and the R2 and MSE remains almost the same.

The most important predictor variables also remain the same and top 20 important variables are mentioned in the below table and the changes due to doubling is high-lighted in yellow.

- *Ridge model:* OpenPorchSF and BsmtQual_EX gains importance over SaleType_Oth and House Style_1.5Unf.
- *Lasso model:* BsmtCond_TA and Condition1 Norm gains importance over Lot Frontage and House Style_2.5Unf.

Ridge Doubled Alpha Co-Efficient	Lasso Doubled Alpha Co-Efficient
Total_sqr_footage	Total_sqr_footage
GarageArea	GarageArea
TotRmsAbvGrd	TotRmsAbvGrd
LotArea	OverallCond
OverallCond	LotArea
LotFrontage	CentralAir_Y
Total_porch_sf	Total_porch_sf
CentralAir_Y	Neighborhood_StoneBr
Neighborhood_StoneBr	BsmtQual_Ex
HouseStyle_2.5Unf	MSSubClass_70
RoofMatl_WdShngl	OpenPorchSF
MSSubClass_70	KitchenQual_Ex
Alley_Pave	Alley_Pave
SaleType_Con	LandContour_HLS
SaleType_ConLD	BsmtCond_TA

KitchenQual_Ex	MasVnrType_Stone
OpenPorchSF	PavedDrive_P
PavedDrive_P	PavedDrive_Y
BsmtQual_Ex	Condition1_Norm
SaleType_CWD	SaleCondition_Partial

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:

I would choose *Lasso regression*. Lasso would perform feature selection and the model will be more robust. For the optimal value of lambda obtained, R2 score and MSE metrics for the model is better for Lasso regression than Ridge regression.

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:

- LotFrontage -> Linear feet of street connected to property
- HouseStyle_2.5Unf -> HouseStyle Two and one-half story: 2nd level unfinished
- Total_porch_sf -> Total Porch Area (Open + Enclosed)
- HouseStyle_2.5Fin -> HouseStyle Two and one-half story: 2nd level finished
- Neighborhood_StoneBr -> Neighborhood Stone Brook

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:

In order for the model to be robust and generalizable, model should be as simple as possible, able to identify the patterns in the data than memorizing them and avoid overfitting. This can be achieved by using regularization method. Robustness can be tested by multiple validation methods on the unseen data. Simpler the model the more the bias but less variance and more generalizable. Use Bias-Variance trade-off. Its implication in terms of accuracy is that a robust and generalizable model will perform equally well on both training and test dataset.