

Answers to 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?

Ridge Regression -

Variable	Coeff(alpha = 0.01)	Coeff(alpha = 0.02)
RoofMatl_Membran	1.014119	0.8964
RoofMatl_WdShngl	1.003659	0.904456
RoofMatl_Metal	0.965965	0.8503
RoofMatl_Roll	0.926958	0.820013
RoofMatl_CompShg	0.92072	0.820946
RoofMatl_Tar&Grv	0.920346	0.817521
RoofMatl_WdShake	0.912477	0.80949
TotalBsmSF	0.216652	0.197964
BsmFinSF1	0.213456	0.195617
GrLivArea	0.172667	0.173191
LotArea	0.165059	0.156858
1stFlrSF	0.148546	0.150996
2ndFlrSF	0.126489	0.12242
OverallQual	0.106425	0.109283
RoofStyle_Shed	0.092021	0.087926

The optimal value of alpha for ridge regression is 0.01. The changes(suppression) in the coefficients when the alpha value is doubled is as shown above in the table. The most important predictor gets changed to RoofMatl_WdShngl after the change is implemented.

Lasso Regression-

Variable	Coeff(alpha=0.0001)	Coeff(alpha=0.0002)
GrLivArea	0.362362	0.324999
OverallQual	0.147967	0.163150
RoofMatl_WdShngl	0.131375	0.085973
GarageCars	0.076611	0.072091
Neighborhood_NoRidge	0.068033	0.070805
BsmFinSF1	0.063536	0.052415
LotArea	0.053121	0.015472

RoofMatl_CompShg	0.041499	0.013852
BsmtExposure_Gd	0.035281	0.037543
Neighborhood_NridgHt	0.03456	0.035478
OverallCond	0.033879	0.024247
Neighborhood_Crawfor	0.032171	0.029852
MSZoning_FV	0.031576	0.020638
Exterior1st_BrkFace	0.029693	0.029007
TotalBsmtSF	0.027916	0.00779

The optimal value of alpha for ridge regression is 0.0001. The changes(suppression) in the coefficients when the alpha value is doubled is as shown above in the table. The most important predictor gets changed to GrLivArea after the change is implemented.

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?

Lasso reduces the number of predictor variables by 49 variables while Ridge reduces only 2 variables very close to zero. In this context the model becomes less complex in Lasso. The variance is also acceptable in the Lasso regression. So its good to go with the Lasso 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?

After the first 5 prominent variables are deleted, below is the list of the prominent variables among the remaining.

Variable	Coefficient
1stFlrSF	0.417031
2ndFlrSF	0.203215
TotalBsmtSF	0.089263
MasVnrArea	0.068567
LotArea	0.055173

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?

To make the model robust and generalisable, we need to be careful about model complexity and should not be either overfitting or underfitting. Here comes the Generalisation methods like Ridge and Lasso regressions. We have seen the same in the results with less variance for Lasso with a little compromise on the model bias . The r squared value got reduced from around 93% in normal regression to 84% in Lasso regression.

The model fits less accurately as the r squared value gets dropped leading to increase in bias but generalising leads to curb in the overfitting which is evident from the working.