Advanced Regression

Assignment Part II - Subjective Questions

Sanjeev S

10/30/22

Master of Science in Machine Learning & Artificial Intelligence

Table of Contents

List of Figures

No table of figures entries found.

List of Tables

No table of figures entries found.

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?

1.1 What is the optimal value of alpha for ridge and lasso regression?

Optimal value of alpha for Ridge Regression is 0.8

Optimal value of alpha for Lasso is 0.0001

1.2 What will be the changes in the model if you choose double the value of alpha for both ridge and lasso?

Changes in Ridge Regression metrics:

- R2 score of train set decreased from 0.947 to 0.92
- R2 score of test set remained same at 0.891

Changes in Lasso metrics:

- R2 score of train set decreased from 0.937 to 0.926
- R2 score of test set decreased from 0.893 to 0.894

1.3 What will be the most important predictor variables after the change is implemented?

```
GrLivArea (log scale)    0.15
1stFlrSF (log scale)     0.11
OverallQual              0.11
LotArea (log scale)      0.09
OverallCond              0.07
MSZoning_RL              0.06
RoofMatl_WdShngl         0.05
MSZoning_FV              0.05
MSZoning_RM              0.05
MSZoning_RH              0.05


GrLivArea (log scale)    0.28
OverallQual              0.14
LotArea (log scale)      0.09
OverallCond              0.08
1stFlrSF (log scale)     0.07
Neighborhood_Somerst     0.03
FullBath_3               0.03
BsmtQual                 0.03
GarageCars_3             0.03
Neighborhood_Crawfor     0.03
```

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?

- It depends on what kind of business problem we are dealing with and use cases.
- If we have a high dimensionality and high correlation in the dataset, then we would want to prefer Lasso regularisation since it penalises less important features more and makes them zero which gives you the benefit of algorithmic feature selection and would make robust predictions.
- Ridge regularisation handles the model complexity by focusing more on the important features which contribute more to the overall error than the less important features.

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?

The 5 most important predictor variables that will be excluded are:

- GrLivArea (log scale)
- OverallQual
- LotArea (log scale)
- OverallCond
- 1stFlrSF (log scale)

The 5 most important predictor variables in the newly designed model are:

- FullBath_3
- MSZoning_RL
- MSZoning_RH
- MSZoning_RM
- MSZoning_FV

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?

4.1 How can you make sure that a model is robust and generalisable?

- A model should not be impacted by outliers in the training data. In simple terms, model performance should remain same even when variation happens in data set.
- A model should be able to predict when new, previously unseen data, drawn from the same distribution as the one used to create the model is provided.
- A model should not overfit on training data, which can affect accuracy in predicting unseen data
- A model should not underfit as well, which fails to identify any relationship between target and predictor variables

In general, model should not be too complex. It should try to bring the variance to constant level by lightly adjusting bias.

4.2 What are the implications of the same for the accuracy of the model and why?

- We decrease variance which will lead to some bias, to make a robust and generalized model. Adding bias will decrease the accuracy of the model.
- Regularization helps in finding an optimal solution between accuracy and robutness to build a better model.