Answers to the 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

- Optimal values for alpha: (ridge = 1000 , Lasso = 0.1). The reason behind this selection is explained in detail in the ipynb file, and hence for the purpose of brevity and respecting the reader's time and patience, it is not duplicated here.
- Changes in the model:
  - Once the value of alpha is doubled, it means that more weight is given to the penalty term of the beta coefficients, and hence more forcing the coefficients towards zero (smaller values), and hence smaller beta values. This can be seen in the table below, that by doubling the alpha, the coefficients have become smaller.
  - Additionally, the R^2 scores have decreased by doubling the alpha, as was already expected from the R^2 curve, since it can be observed that by increasing the value of alpha, the R^2 score has decreased in the R^2 curve. This means the model is becoming slightly under-fit.
- Regarding the most important predictors:
  - Ridge: in the ridge regression, the first 5 predictors are the same for the optimal value of alpha, and for the doubled value, however, their ranking is different.
  - Lasso: in the lasso regression, just the first 2 variables are the same (with smaller coefficients), however, the rest of the most important variables are different from the optimal alpha case, however, with zero (or very small coefficients).

	Alpha	R^2 score	First 5 significant variables
Ridge	1000 (optimal)	On train set: 0.6984 On test set: 0.7040	Feature / Coefficient 0 OverallQual 0.293827 3 1stFlrSF 0.169078 4 2ndFlrSF 0.147993 2 TotalBsmtSF 0.143683 1 BsmtFinSF1 0.089075

	2000 (double)	On train set: 0.6024 On test set: 0.6103	**Feature** / **Coefficient** **0** OverallQual 0.219880 **3** 1stFlrSF 0.133906 **2** TotalBsmtSF 0.124253 **4** 2ndFlrSF 0.103909 **1** BsmtFinSF1 0.074101
Lasso	0.1 (optimal)	On train set: 0.7266 On test set: 0.7348	**Feature** / **Coefficient** **0** OverallQual 0.545742 **3** 1stFlrSF 0.227597 **4** 2ndFlrSF 0.102673 **1** BsmtFinSF1 0.042236 **2** TotalBsmtSF 0.031533
	0.2 (double)	On train set: 0.6505 On test set: 0.6638	**Feature** / **Coefficient** **0** OverallQual 0.532373 **3** 1stFlrSF 0.151479 **35** CBlock -0.000000 **26** WdShake 0.000000 **27** WdShngl 0.000000

Next questions continue on the next page.

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

Coefficients (beta values) of ridge and lasso:

```
lrm_lasso.coef_
```

```
array([ 0.54574164,  0.04223552,  0.03153277,  0.22759739,  0.10267332,
       -0.        , -0.        ,  0.        , -0.        ,  0.        ,
       -0.        ,  0.        ,  0.        ,  0.        ,  0.        ,
       -0.        ,  0.        ,  0.        ,  0.        , -0.        ,
       -0.        ,  0.        ,  0.        ,  0.        , -0.        ,
        0.        , -0.        ,  0.        , -0.        , -0.        ,
       -0.        ,  0.        ,  0.        , -0.        , -0.        ,
       -0.        , -0.        ,  0.        ,  0.        , -0.        ,
       -0.        , -0.        , -0.        ,  0.        ,  0.        ,
        0.        ,  0.        ])
```

```
lrm_ridge.coef_
```

```
array([ 2.93827048e-01,  8.90754677e-02,  1.43683110e-01,  1.69077724e-01,
        1.47992543e-01,  3.48551295e-04, -1.23370300e-02,  1.33105859e-02,
       -2.13463737e-03,  4.93280243e-02, -5.48859459e-02,  2.28122807e-02,
        4.61771475e-02,  9.15677598e-03,  2.27671475e-03, -3.86197330e-03,
       -3.72168874e-05,  3.51633647e-03,  9.70785196e-04, -1.23370300e-02,
       -9.99535970e-04,  2.08813903e-03,  7.07504659e-04,  2.39350310e-04,
       -2.94668748e-04,  1.39622822e-03,  9.16712028e-04,  7.08487110e-04,
       -4.65445734e-04, -4.65445734e-04, -6.19130022e-02,  1.79982972e-02,
        1.90195967e-04, -1.80875043e-03, -4.65445734e-04, -4.65445734e-04,
       -6.19130022e-02,  1.96209702e-02,  8.05297455e-04, -4.26416957e-04,
       -1.33732390e-03, -3.59103728e-03, -6.89667023e-04,  1.98570453e-05,
        7.44104346e-04,  5.21100643e-02,  1.40346479e-03])
```

Comparison of the R^2 scores:

	Base	Ridge	Lasso
train set	0.868659	0.698468	0.726642
test set	0.785986	0.704066	0.734876

As can be seen, by comparing the values of R^2 score of different models, the Lasso model is chosen due to the following reasons:

- The higher R^2 score compared to the Ridge model, and the lower difference between R^2 score of train and test in comparison to the base model. In other words, the R^2 score of the base model is higher than the Lasso, however, the difference between the R^2 score of train and test sets of the base model is higher than that of the Lasso model, which means the base model is slightly over-fit.
- The Lasso model has only 5 non-zero coefficients for the features, and while having a higher R^2 score compared to the Ridge, seems to be simpler and more generalisable.

Next questions continue on the next page.

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

Apologies, but this has never happened since lasso works with the available features and just tries to find the coefficients. This case has not even happened in the exercises demonstrated for us in the lectures. However, I am keen to learn more on this and see why this might happen.

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

When I see the difference between the R^2 score of train set and the R^2 score of test set are quite small, then I would conclude that the model is not under-fit nor over-fit, and hence is robust and generalisable. This also shows that the model has learnt the patterns in the data, and hence its performance on the test (unseen) data will be similar to its performance on the train data.

The accuracy of a robust and generalisable model is smaller than the over-fit model, however, it performs much better on the unseen (test) data compared to the over-fit model.

The reason for having smaller R^2 score (accuracy) is that the robust model has more mis-classifications (or predictions with higher errors) compared to an over-fit model with a higher R^2 score (accuracy) which has been forced to classify/predict more data points in the training set. However, the over-fit model just learns the data points and not the patterns within them, whereas a more robust/generalisable model have understood the patterns in the data even if it might have more mis-classifications/predictions with higher errors.