Module 2 Assignment on Linear Regression - 2 - V1

Jingwen Zhong // Graduate Student

2/24/2021

Module Assignment Questions

In this assignment, you will use the `Auto` data set with 7 variables (one response `mpg` and six numerical) and $n = 392$ vehicles. For sake of simplicity, categorical variables were excluded. Before each randomization used, use `set.seed(99)` so the test results are comparable.

Q1) (***Forward and Backward Selection***)

In `Module 1 Assignment`, `Q2`, you fitted `Model 3` with `mpg` as the response and the six numerical variables as predictors. This question involves the use of `forward` and `backward` selection methods on the same data set.

a. Using `OLS`, fit the model with all predictors on `mpg`. Report the predictors' coefficient estimates, R^2_{adj}, and MSE. Note: The method in `lm()` is called ordinary least squares (OLS).

Table 1: Report of Model_Full fitted by OLS

	coefficient of cylinders	displacement	horsepower	weight	acceleration	year	R^2adj	MSE
OLS model	-0.3299	0.0077	-4e-04	-0.0068	0.0853	0.7534	0.8063	11.8009

---

b. Using `forward selection method` from `regsubsets()` and `method="forward"`, fit MLR models and select the `best` subset of predictors. Report the best model obtained from the default setting by including the predictors' coefficient estimates, R^2_{adj}, and MSE.

Table 2: Report of Best Model fitted by forward selection method

	Coefficient of weight	Coefficient of year	R^2adj	MSE
forward selection Best_Model	-0.0066	0.7573	0.8072	11.7454

---

c. What criterion had been employed to find the best subset? What other criteria exist? Explain.

I used BIC to find the best subset, BIC is derived from a Bayesian, and also is estimates of test MSE.

We can also use C_p, AIC, and adjusted R^2 to find the best models. C_p is estimates of test MSE too, it adds a penalty to the training SSE in order to adjust for the fact that the training error tends to underestimate the test error; AIC is defined for a large class of models fit by maximum likelihood. for C_p, AIC and BIC, the smaller value is, the lower test error the model has. However, a large value of adjusted R2 indicates a model with a small test error

---

d. Using `backward selection method` from `regsubsets()` and `method="backward"`, fit MLR models and select the `best` subset of predictors. Report the best model obtained from the default setting by including predictors, their coefficient estimates, R^2_{adj}, and MSE.

Table 3: Report of Best Model fitted by backward selection method

	Coefficient of weight	Coefficient of year	R^2adj	MSE
forward selection Best_Model	-0.0066	0.7573	0.8072	11.7454

---

e. Compare the results obtained from `OLS`, `forward` and `backward` selection methods (parts a, b and d): What changed? Which one(s) is better? Comment and justify.

Compared The model we used in OLS method to The best model of forward and backward selection methods, everything changed.

In OLS method, we use full model which contains 6 predictors. The best models of forward and backward selection methods are the same, they both contains 2 predictors which are weight and year.

The absolute value of coefficients of Weights is a little smaller in the best models of forward and backward selection methods which means weights become a little less important, and the absolute value of year is larger, which means when year increase 1, mpg increase more.

R_a^2dj and MSE changed too. R_a^2dj become larger, and MSE become smaller, which makes the best models of forward and backward selection methods better than the full model fitted by OLS.

Q2) (*Cross-Validated with k-Fold*)

What changes in model selection results and the coefficient estimates when cross-validated set approach is employed? Specifically, we will use k-fold cross-validation (`k-fold CV`) here.

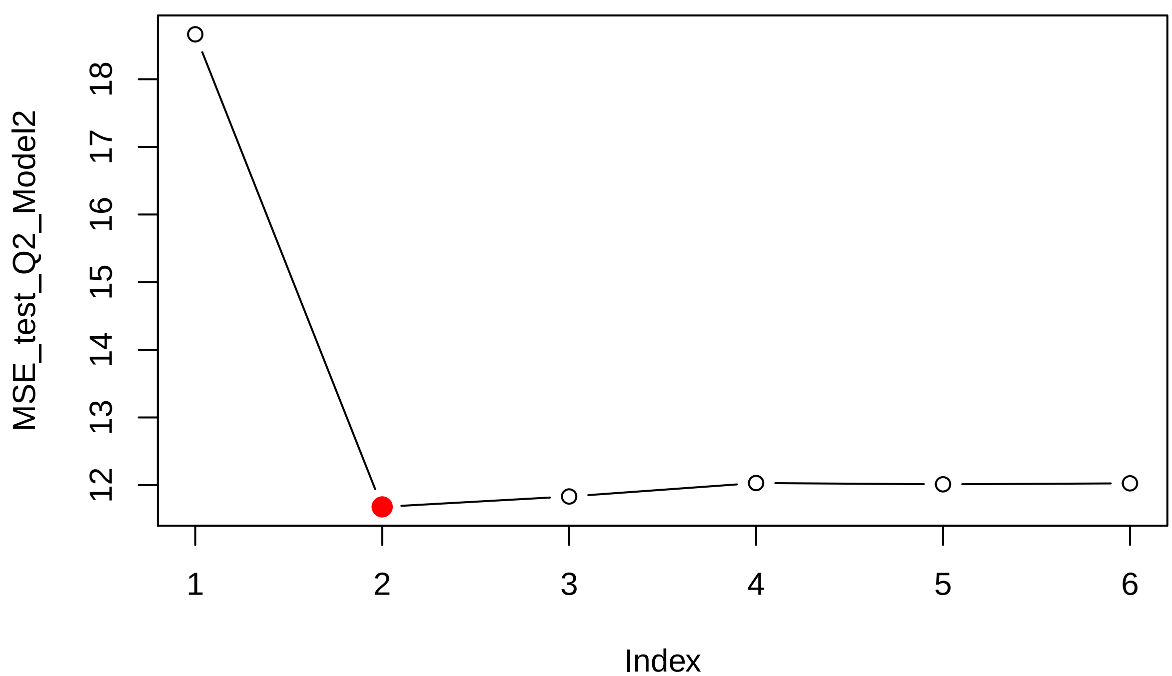
a. Using the 5-fold CV approach, fit the OLS MLR model on `mpg` including all the predictors (don't use any subset selection). Report all the predictors' coefficient estimates in the OLS model (using all folds), the averaged MSE_{train}, and the averaged MSE_{test}.

Table 4: Full model fitted by OSL using 5-fold cv

	coefficient of cylinders	displacement	horsepower	weight	acceleration	year	MSE_train	MSE_test
fold 1	-0.4761	0.0064	-0.0063	-0.0063	-0.0967	0.7371	10.5289	16.5448
fold 2	-0.5374	0.0112	0.0009	-0.0067	0.1039	0.7540	11.4086	12.3383
fold 3	-0.0831	0.0046	0.0098	-0.0073	0.1627	0.7719	11.6421	11.6155
fold 4	-0.0666	0.0034	-0.0012	-0.0068	0.1325	0.7647	12.2663	8.7794
fold 5	-0.4404	0.0116	-0.0069	-0.0068	0.0983	0.7430	11.7909	10.8522

---

b. Using the 5-fold CV approach and `forward selection method`, fit MLR models on `mpg` and select the `best` subset of predictors. Report the best model obtained from the default setting by including the predictors' coefficient estimates (this depends on what predictors you keep in the model), the averaged MSE_{train}, and the averaged MSE_{test}.



```
## The coefficient for best model is

##       weight         year 
## -0.006632075  0.757318281
```

Table 5: forward: 5 fold's Average MSE for different number of predictors

	MSE_train	MSE_test
1 p	18.6568	18.6636
2 p	11.6367	11.6780
3 p	11.5843	11.8315
4 p	11.5563	12.0303
5 p	11.5304	12.0116
6 p	11.5274	12.0260

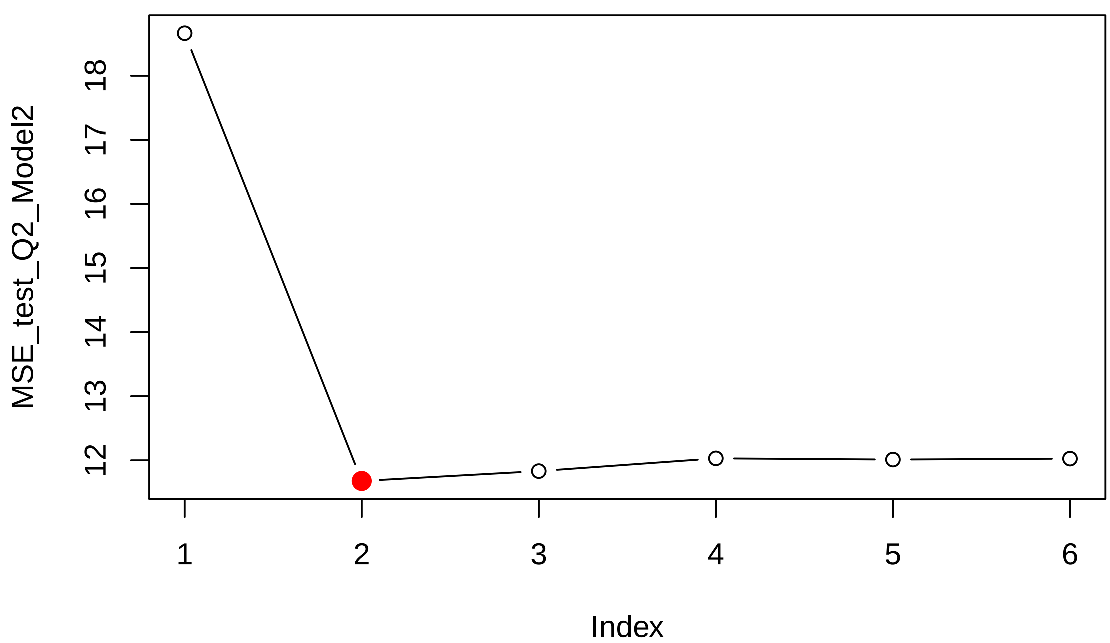
---

c. Compare the MSE_{test}'s. Explain.

MSE_test of the best model fitted by forward selection using k-fold cv is smaller than the Full model fitted by OSL using k-fold cv. We use MSE_test to decide which model is better, and the smaller MSE_test the better the model is, so best model fitted by forward selection is better.

---

d. Using the 5-fold CV approach and `backward selection method`, fit MLR models on `mpg` and select the **best** subset of predictors. Report the best model obtained from the default setting by including the predictors' coefficient estimates, the averaged MSE_{train}, MSE_{test}.




```
## The coefficient for best model is

##   (Intercept)        weight          year
## -14.347253018  -0.006632075   0.757318281
```

Table 6: backwards: 5 fold's Average MSE for different number of predictors

	MSE_train	MSE_test
1 p	18.6568	18.6636
2 p	11.6367	11.6780
3 p	11.5843	11.8315
4 p	11.5563	12.0303
5 p	11.5304	12.0116
6 p	11.5274	12.0260

---

e. Did you come up with a different model on parts b and d? Are the predictors and their coefficient estimates same? Compare and explain.

I did not come up with a different model on part b and d, the predictors and their coefficient estimates are the same. We use the forward and backward selection with k-fold cv to find the minimal avarage MSE of testing data set, and in both models, the MSE of the model with 2 predictors(weights and year) is the smallest. Since we choose the same model, the predictors and their coefficient estimates should be the same.

f. Which fitted model is better among parts a, b, and d? Why? Justify.

The model with 2 predictors(weights and year) is better, because it has smaller MSE_test

Q3) (*Shrinkage Methods*)

Results for `OLS`, `lasso`, and `ridge` regression methods can be comparable. Now, you are expected to observe that ridge and lasso regression methods may reduce some coefficients to zero (so in this way, these features are eliminated) and shrink coefficients of other variables to low values.

In this exercise, you will analyze theses estimation and prediction methods (OLS, ridge, lasso) on the `mpg` in the Auto data set using $k-fold$ cross-validation test approach.

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1
```

a. Fit a ridge regression model on the entire data set (including all six predictors, don't use yet any validation approach), with the optimal λ chosen by `cv.glmnet()`. Report $\hat{\lambda}$, the predictors' coefficient estimates, and MSE.

Table 7: Full model fitted by ridge regression

	Best_lambda	coefficient of cylinders	displacement	horsepower	weight	acceleration	year	MSE
ridge regression model	0.6487	-0.4552	-0.007	-0.0222	-0.004	-0.0689	0.6602	12.1544

---

b. Fit a lasso regression model on the entire data set (including all six predictors, don't use yet any validation approach), with the optimal λ chosen by `cv.glmnet()`. Report $\hat{\lambda}$, the predictors' coefficient estimates, and MSE.

Table 8: Full model fitted by lasso regression

	Best_lambda	coefficient of cylinders	displacement	horsepower	weight	acceleration	year	MSE
lasso regression	0.0389	-0.0906	0	0	-0.0064	0.0499	0.7394	11.6272

---

c. Compare the parts a and b in Q3 to part a in Q1. What changed? Comment.

Everything changed. MSE: ridge>OLS>lasso, which make the model fitted by lasso regression a better choice. displacement and horsepower in the model fitted by lasso becomes 0 which means displacement and horsepower are eliminated.The acceleration in ridge model become negative.

Both Ridge and lasso regression are very similar to least squares, except that the coefficients ridge and lasso are estimated by minimizing a slightly different quantity, and lead the coefficient estimates can be shrunk towards zero. We can see that in ridge regression model, no coefficient of predictors equals to 0, it because The penalty $\lambda \sum \beta_j^2$ in ridge regression will shrink all of the coefficients towards zero, but it will not set any of them exactly to zero (unless $\lambda = \infty$).

---

d. How accurately can we predict `mpg`? Using the three methods (OLS, ridge and lasso) with all predictors, you will fit and test using 5-fold cross-validation approach with the optimal λ chosen by `cv.glmnet()`. For each, report the averaged train and test errors (MSE_{train}, MSE_{test}):

1) Fit an `OLS` model.
2) Fit a `ridge` regression model.
3) Fit a `lasso` regression model.

```
##  MSE_Train and MSE_test of OLS model is: 11.52738 and 12.02604 and
##  average accuracy is: 0

##  MSE_Train and MSE_test of Ridge regression model is: 12.09278 and 12.49224 and
##  average accuracy is: 0

##  MSE_Train and MSE_test of Lasso Regression model is: 11.57724 and 11.90807 and
##  average accuracy is: 0
```

---

e. Write an overall report on part d by addressing the inquiry, `how accurately can we predict mpg?`. Is there much difference among the test errors resulting from these three approaches? Show your comprehension.

By looking at the MSE of test set, Lasso Regression performs well because it has smaller MSE of test set, it means the error between the real values and the predict values is smaller. By calculating how many values are predicted exactly right in these 3 models(see average accuracy in part d), turns out there are no value is predicted exactly right. I think we can only discuss how close the predicted values are to the real value by calculating the MSE.

Besides, there are predictors that have coefficients that are equal zero in lasso regression model, so lasso regression can perform better in this situation. However, the results of these 3 models are not too much different, one reason might be our training set is not large enough. Also, by increasing k to 10, I got larger MSE in all these 3 models(not showing here).

Table 9: Reports of MSE_train and MSE_test on 3 models

	MSE_train	MSE_test
OLS	11.5274	12.0260
Ridge	12.0928	12.4922
Lasso	11.5772	11.9081

f. (BONUS) Propose a different model (or set of models) that seem to perform well on this data set, and justify your answer.

I choose Principal Components Regression, and find the MSE of this model is bigger than other models in previous questions.

```
##
## Attaching package: 'pls'
```

```
## The following object is masked from 'package:stats':
##
##     loadings
```

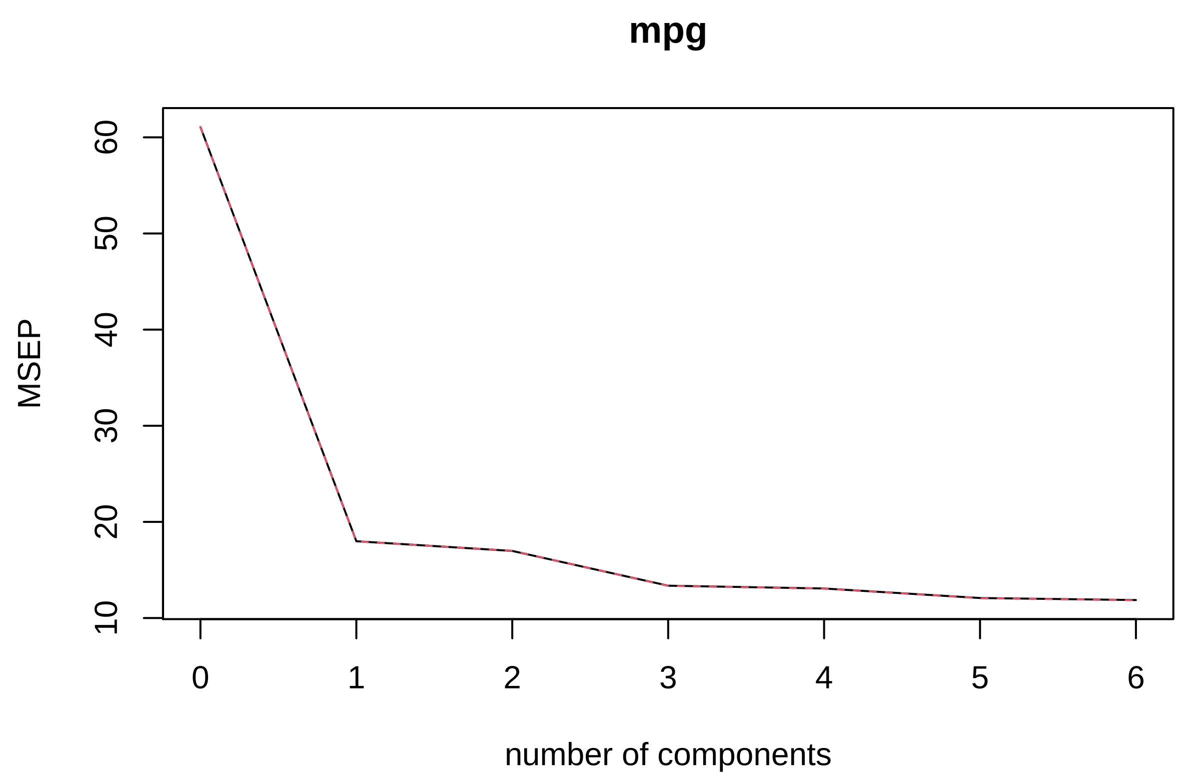


Table 10: Full model fitted by Principal Components Regression

	coefficient of cylinders	displacement	horsepower	weight	acceleration	year	MSE
Principal Components Regression full model	-0.5627	0.8035	-0.0151	-5.7714	0.2353	2.7752	13.9858

---

g. (BONUS) Include categorical variables to the models you built in part d, Q3. Report.

Table 11: Reports of MSE_train and MSE_test on 3 models

	MSE_train	MSE_test
OLS	11.5274	12.0260
Ridge	11.3933	11.8649
Lasso	10.8697	11.3538

h. (GOLDEN BONUS) Propose a model (or set of models) that seem to perform well on this data set, and justify your answer. Make sure that you are evaluating model performance using 5-fold cross-validation approach. You can transform the data, scale and try any methods. When MSE_{test} is the lowest (under the setting of Q3, part d) in the class, your HW assignment score will be 100% (20 pts).

i. (BONUS) You can make a hybrid design in model selection using all the methods here in a way that yields better results. Show your work, justify and obtain better results in part d, Q3.

I hereby write and submit my solutions without violating the academic honesty and integrity. If not, I accept the consequences.

How long did the assignment work take?: 10 hrs

References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.