part1

Cui Qingxuan

2024-12-01

Contents

1 Assignment 1

1.1 Quality of fit, prediction and model with linear regression.

```
## The MSE of training data is: 0.005709117
```

```
## The MSE of test data is: 722.4294
```

The model performs well on the training data (MSE = 0.0057), but has a large error on the test data (MSE = 722.4294), indicating that the model cannot be generalized and there is a serious overfitting phenomenon.

1.2 Report the cost function

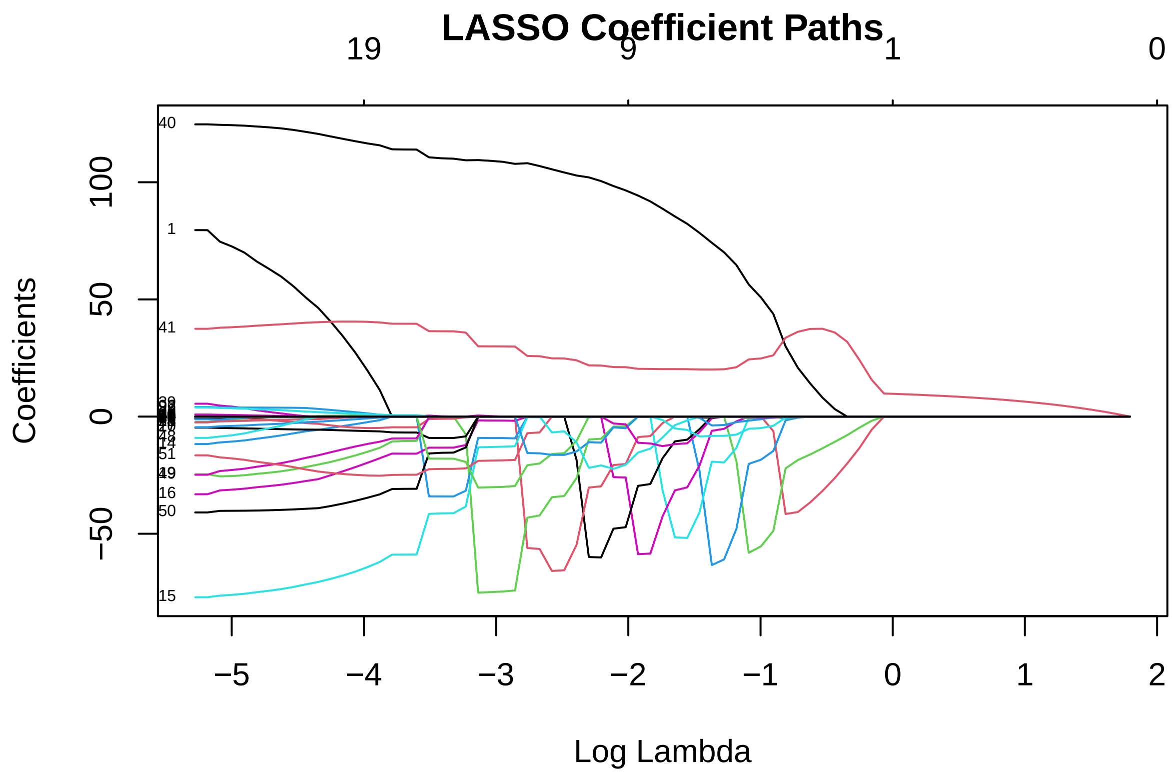
The cost function for **LASSO regression** is defined as:

$J(\beta) = \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \mathbf{X}_i \cdot \beta\right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$

Where lambda controls the regularization strength.

1.3 interpret LASSO Coefficient Paths Plot

```
##
## Coefficients for different lambda values:
```



1. This graph shows the path that the regression coefficient of each feature in the **LASSO regression model** changes with the regularization parameter $\log(\lambda)$: the horizontal axis is $\log(\lambda)$, and the vertical axis is the value of the regression coefficient for each feature.
2. With the increase of λ (from right to left), the regularization force is enhanced, and the coefficients of most features are compressed to 0, leaving only a few important features with non-zero coefficients, which realizes feature selection.

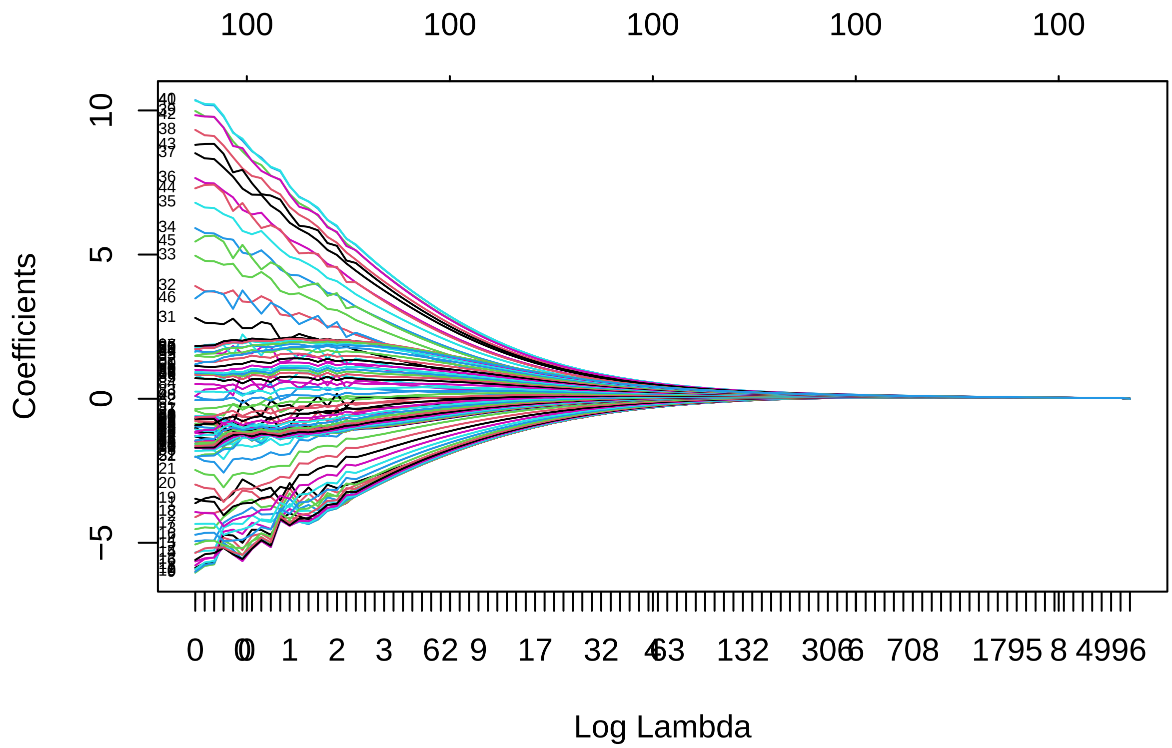
1.4 model with only three features

```
## Lambda with 3 non-zero coefficients: 0.8530452
```

```
## The remaining features are:
```

```
##  Channel6  Channel7 Channel41
## -5.580149 -1.793787 15.696612
```

1.5 fit Ridge regression and compare the plots from steps 3 and 4



LASSO coefficient path diagram:

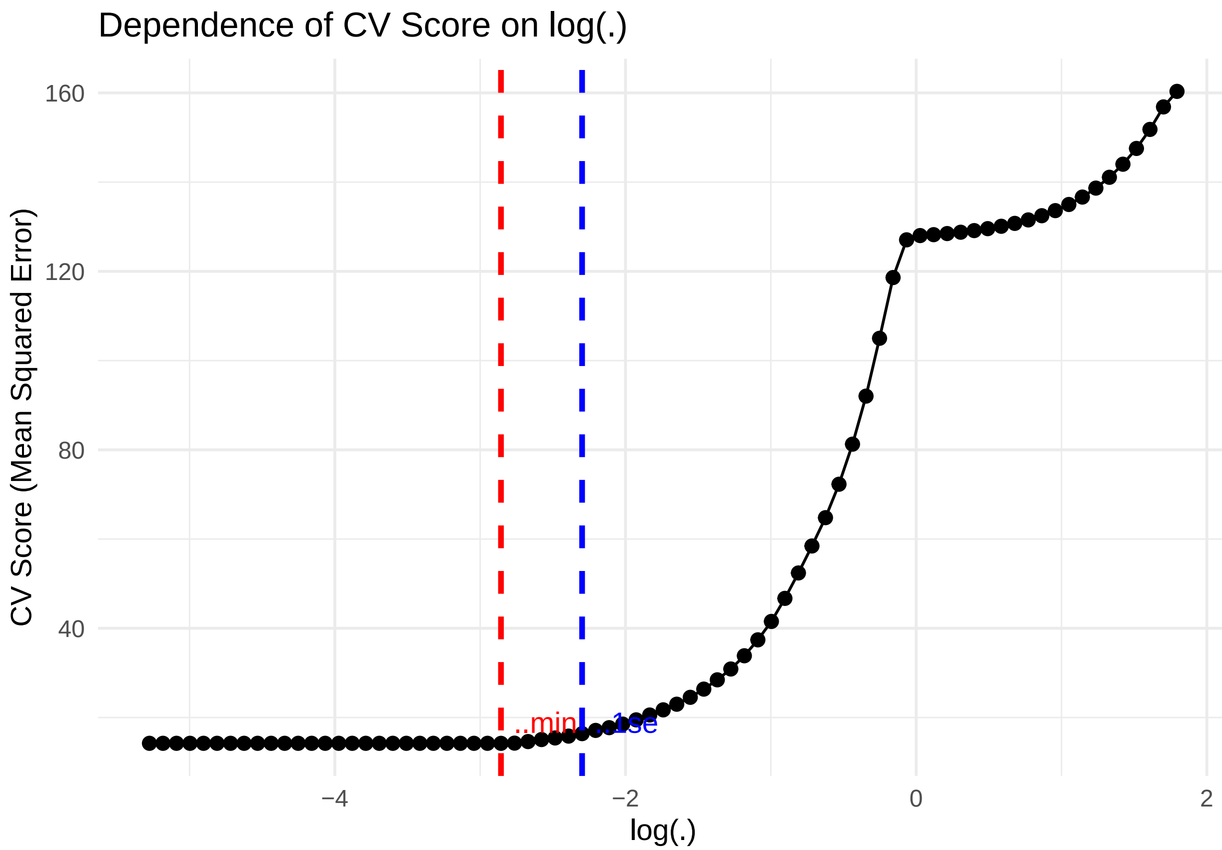
As λ increases, the coefficients of many features are compressed to 0, and the model becomes sparse.

Ridge coefficient path map:

As λ increases, the coefficients of all features gradually shrink, but they do not become zero.

1.6 Task 5

1.6.1 dependence of the CV score on $\log(\lambda)$



(comment)

1.6.2 Optimal *lambda*

```
## the optimal lambda:  0.05744535
```

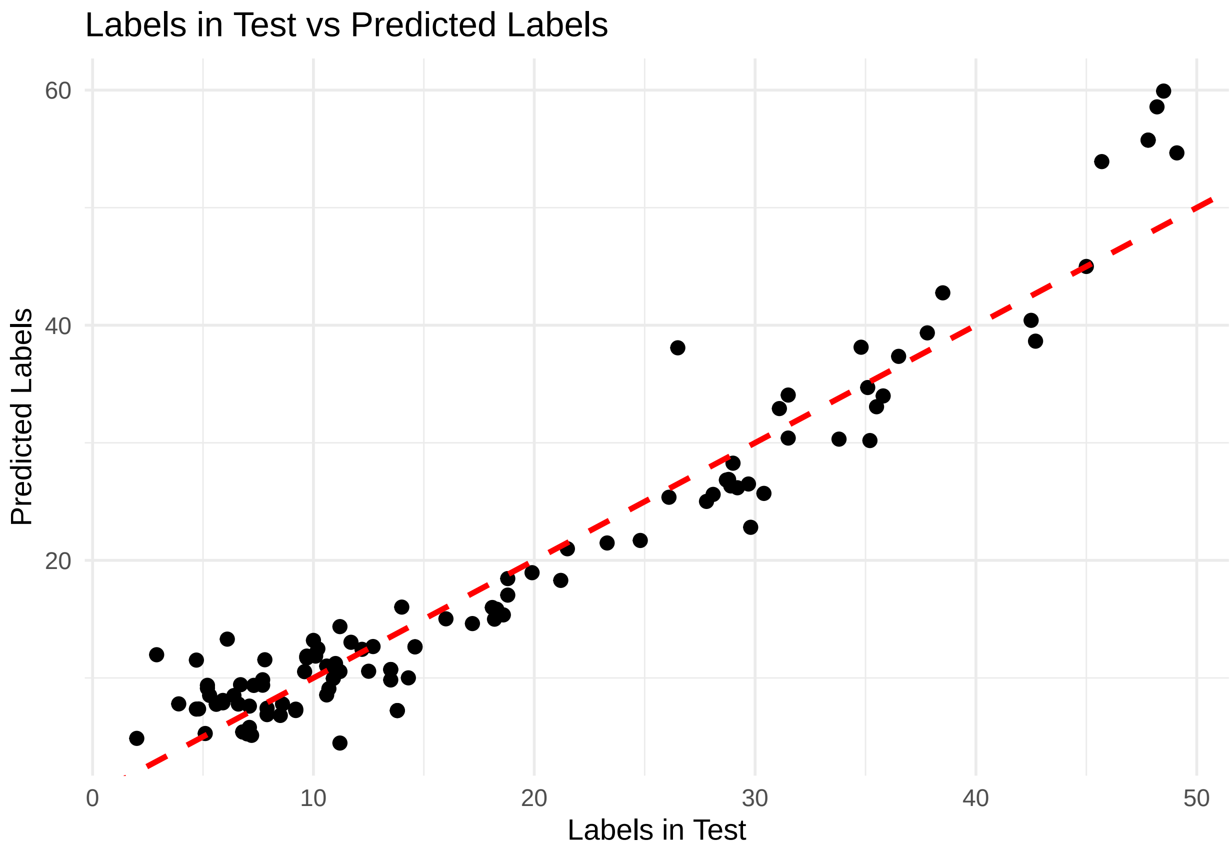
```
## the number of variables:  8
```

```
## log_lambda = -4, mse =  13.48375
```

```
## lambda =  0.05744535 mse =  13.67339
```

The difference in MSE values is very small, approximately 0.19, which is unlikely to be statistically significant.

1.6.3 Scatter plot of the original test versus predicted test values



The model performs well overall, as most predictions are close to the actual values.

The scatterplot shows that the model captures the general trend of the data, with no significant systematic errors.

But a few outliers at higher values suggest the model may struggle with extreme cases or higher variability in predictions for larger labels.