## 5)Lasso & Ridge Regularization

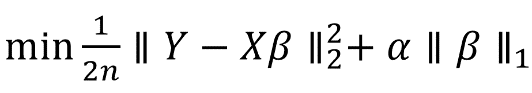
In machine learning, over fitting is likely to happens when a model is excessively complicated. Under this situation, the machine only memorizes the training data, instead of learning to predict the data from given features. In our problem, we have only about 500 tuple of data with 20 different features. Thus, the chance of our model to memorize data from these features is relatively high.

Before, we use cross validation to limit the affection of over fitting during the assessing of a model. In this way, we can only determine if a given model is good or not. Instead, if we want to generate better learning models with little over fitting, we may use method as regularization.

In regularization, we introduce a regularization term or so called loss function, which is usually a penalty on the complexity of our model. Therefore, it can improve the generalization ability of our model.

###### Lasso Regularization

In Lasso regularization, the regularization term is considering the -norm of the regression coefficient vector. And the penalty function has the following form:



For different extent of penalty on the complexity, we can choose various value of α. And in our solution, we choose it in the range {0.1,0.01,0.001} as required. The corresponding performance of the regression model is shown in Figure 1, from which we can see that the RMSE is smaller with a little value of α.

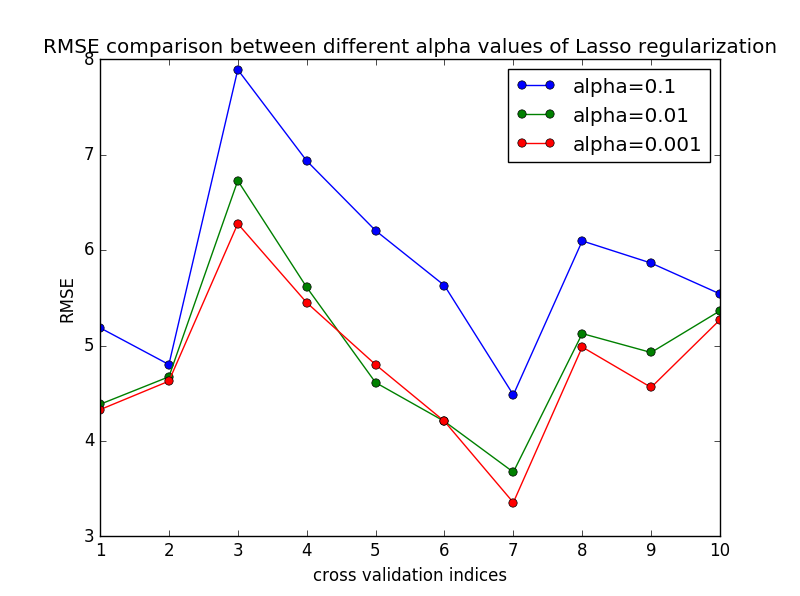
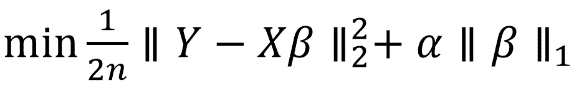


Figure 1 RMSEs of different alpha values in Lasso regularization

After further research, we know that the result of linear regression almost overlaps the result of α=0.001, which means that the RMSE of pure linear regression is smaller than those of Lasso regression. Thus, in this problem, the control of over fitting does not bring a better model to us. However, this does not mean it is useless. After checking the result of the regression coefficients, we are shocked by its sparsity. When α=0.1, only 3 coefficients are nonzero and all the other 17 are 0. Thus, in this way, we have found a great sparse solution of our problem.

###### Ridge Regularization

In Ridge regularization, the regularization term is considering the -norm of regression coefficient vector. And the penalty function has the following form:



Similar as Lasso regularization, we set the value of α in the range {0.1,0.01,0.001} and the performance of our regression models are shown in Figure 2. This time, the curves of RMSE for different α are very close to each other.

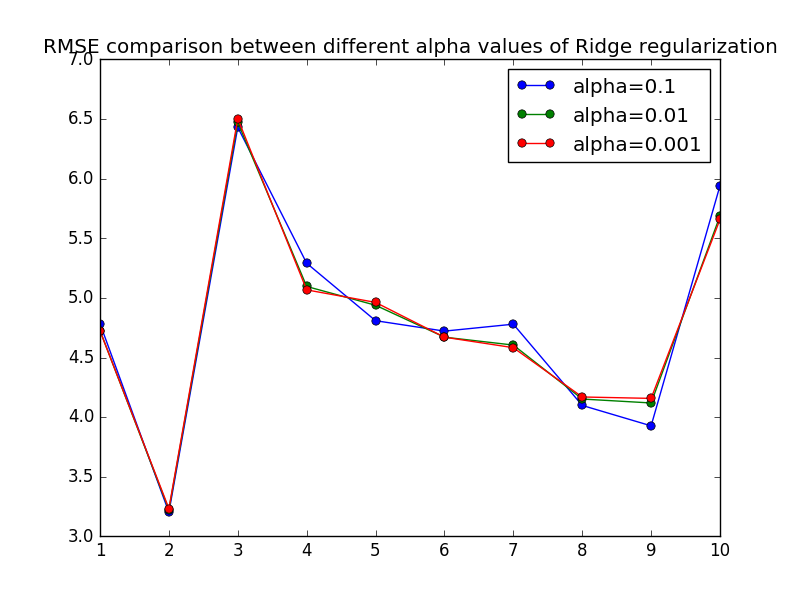


Figure 2 RMSEs of different alpha values in Ridge regularization

For a clearer comparison of the all RMSE values, we calculate the mean RMSE upon every 10-fold cross validation and we put all of them in Table 1:

Table 1 mean RMSE values under different settings

|  |  |  |
| --- | --- | --- |
| Value of α | Lasso regularization | Ridge regularization |
| 0.1 | 5.8639 | 4.7981 |
| 0.01 | 4.9307 | 4.7686 |
| 0.001 | 4.7861 | 4.7725 |

From the above table, we find out that for Lasso regularization, the RMSE is worsened but the sparsity of regression coefficients is obtained. And for Ridge regularization, we get the best RMSE result at α=0.01, which prove that with the control of over fitting, we may get a better model.