Kevin Kyunggeun Jang

ANLY-590

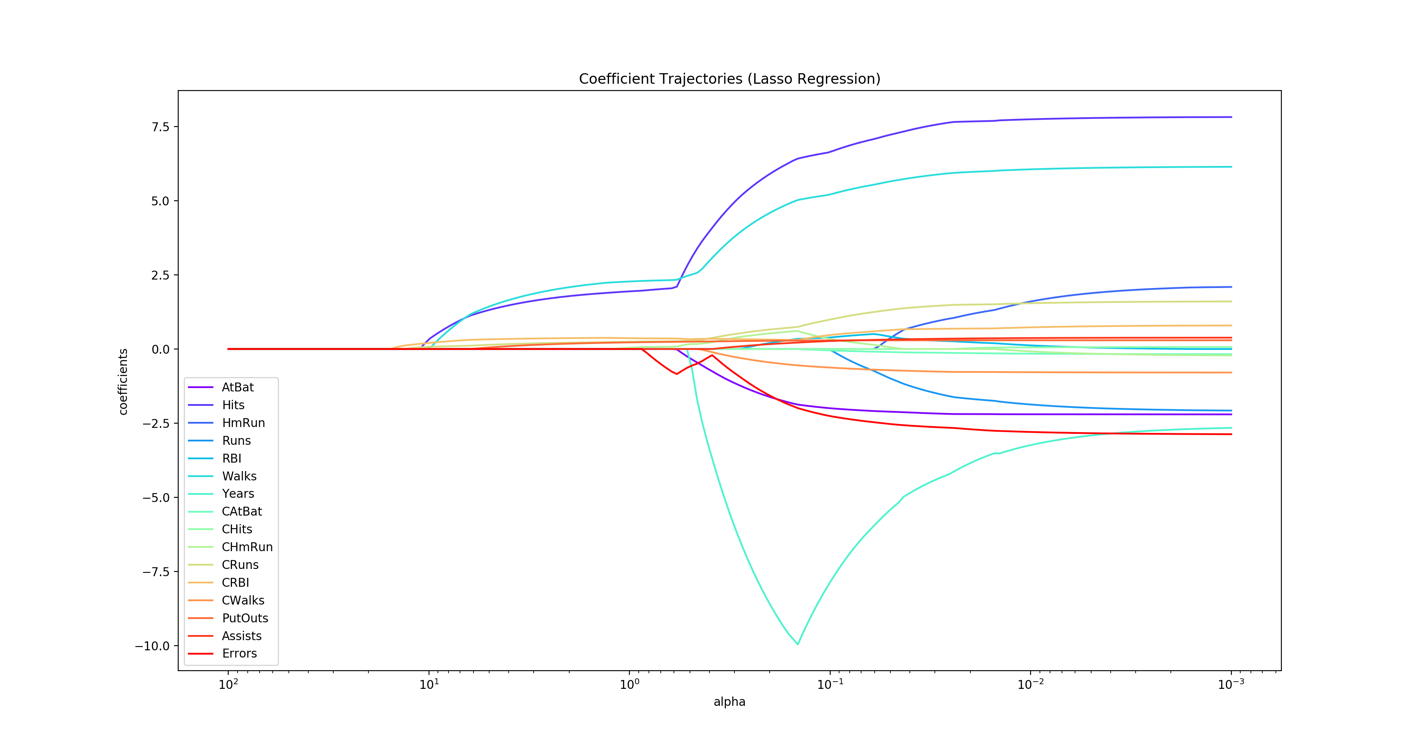
Homework 0

09/19/18

1. **Regularization**. Using the accompanying *Hitters* dataset, we will explore regression models to predict a player's Salary from other variables. You can use any programming languages or frameworks that you wish.

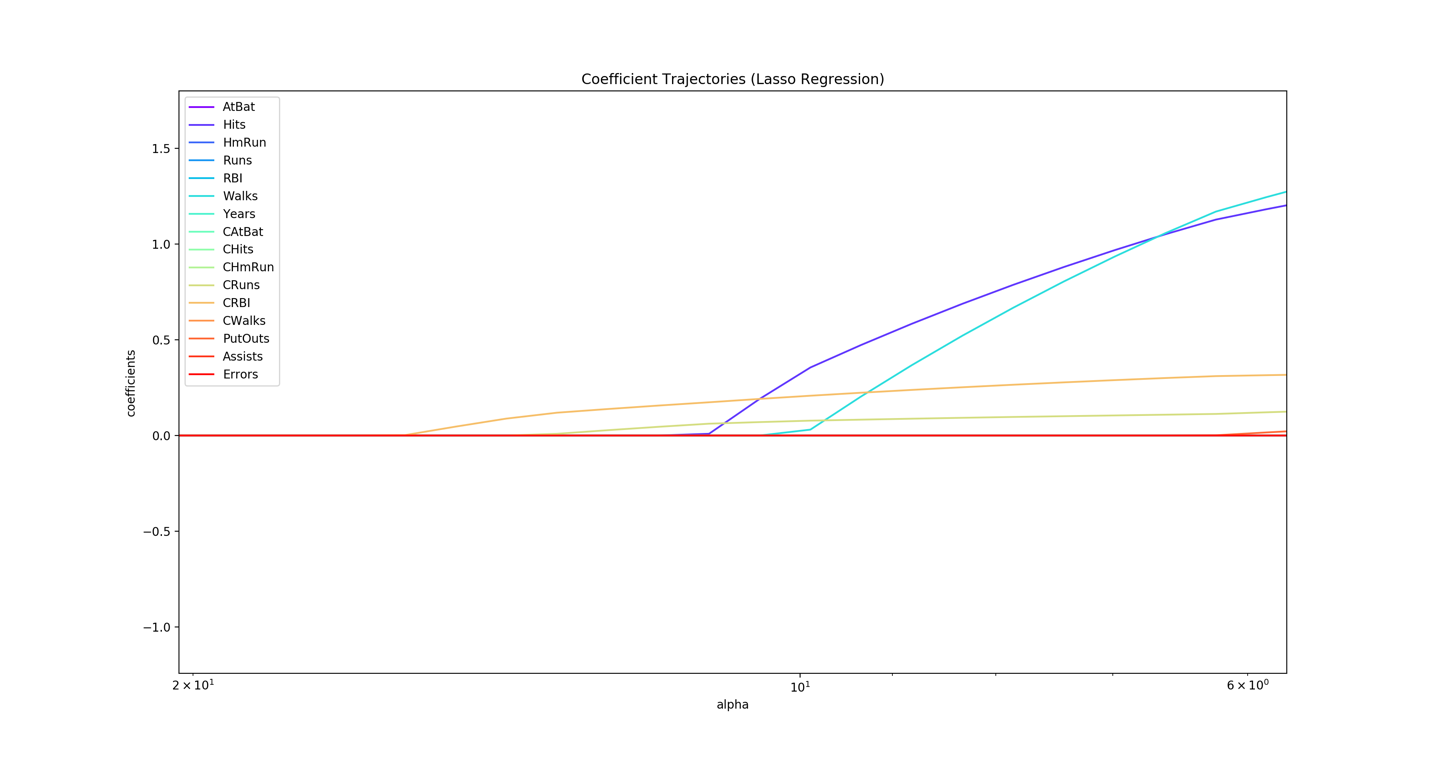
1.1 Use LASSO regression to predict Salary from the other numeric predictors (you should omit the categorical predictors).

- Create a visualization of the coefficient trajectories.



(**Figure 1**. Lasso Regression Coefficient Trajectories)

- Comment on which are the final three predictors that remain in the model.



(**Figure 2**. Zoomed Lasso Regression Coefficient Trajectories)

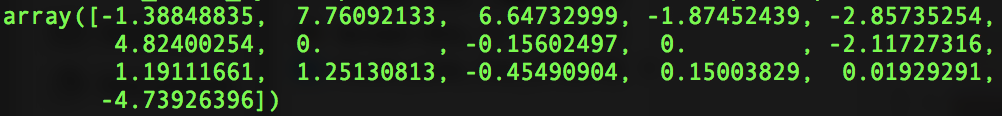
- Figure 2 is the zoomed image of the left part of the Figure 1. As one can see from the above figure, the final three numeric predicators that are still remaining in the model are:

- Use cross-validation to find the optimal value of the regularization penality.

- ‘LassoCV’ function from ‘sklearn’ module was used in python to get this value and it turned out that the optimal alpha value is:

- How many predictors are left in that model?

- To answer this question, a new Lasso model was created with the optimal alpha value (0.0635481759985745) that was achieved from the previous question. With the new Lasso model, coefficients were printed out to the console to count the number of predicators with the non-zero coefficient. Here is how the array of coefficients look like:

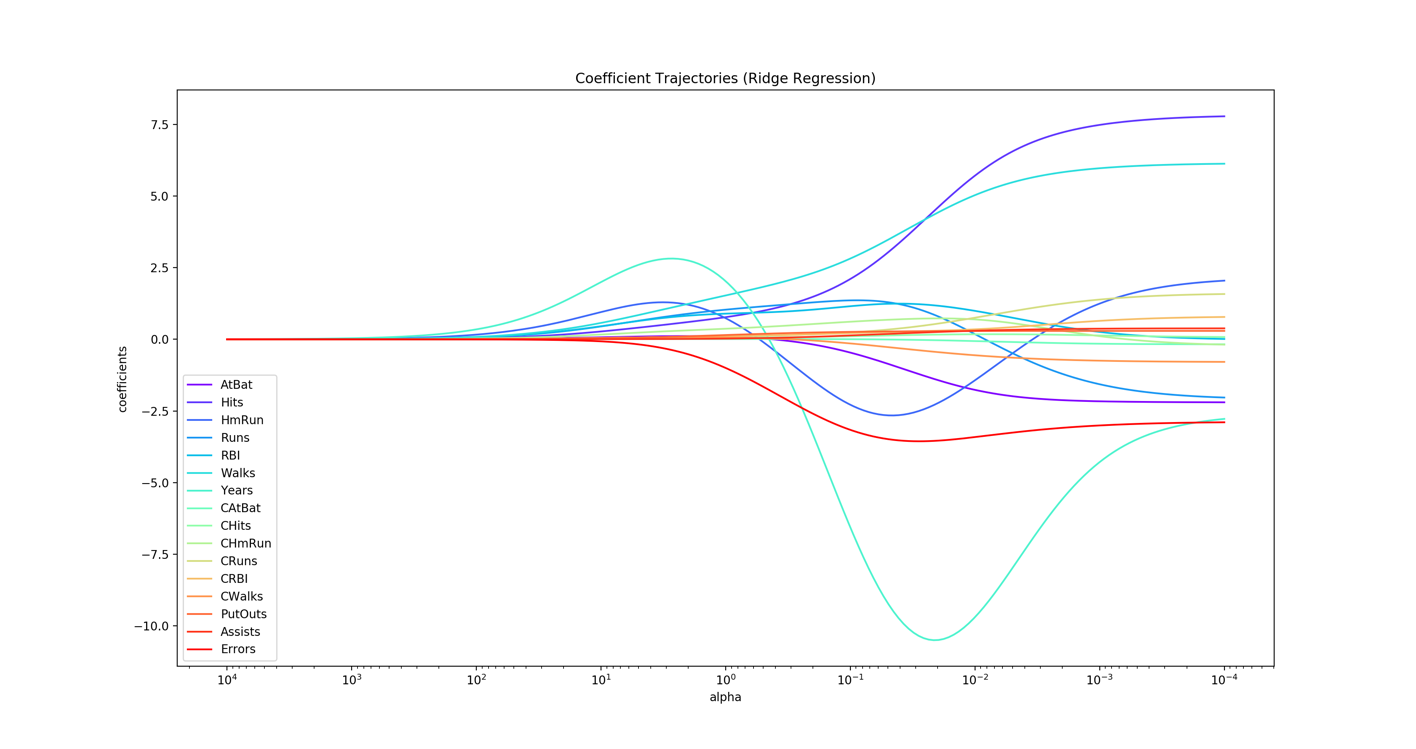


(**Figure 3**. Array of Optimal Lasso Model Coefficients)

- As one can see from Figure 3, out of sixteen predicators, there are two predicators with zero coefficients and fourteen predicators with non-zero coefficient values. Therefore, we can conclude that there are total of fourteen predicators left in the model.

1.2 Repeat with Ridge Regression.

- Visualize coeffecient trajectories.



(**Figure 4**. Ridge Regression Coefficient Trajectories)

- Use cross-validation to find the optimal value of the regularization penalty.

- ‘RidgeCV’ function from ‘sklearn’ module was used in python to get this value and it turned out that the optimal alpha value is:

1. **Short Answer.**

* Explain in your own words the bias-variance tradeoff.
* What role does regularization play in this tradeoff?
* Make reference to your findings in number (1) to describe models of high/low bias and variance.