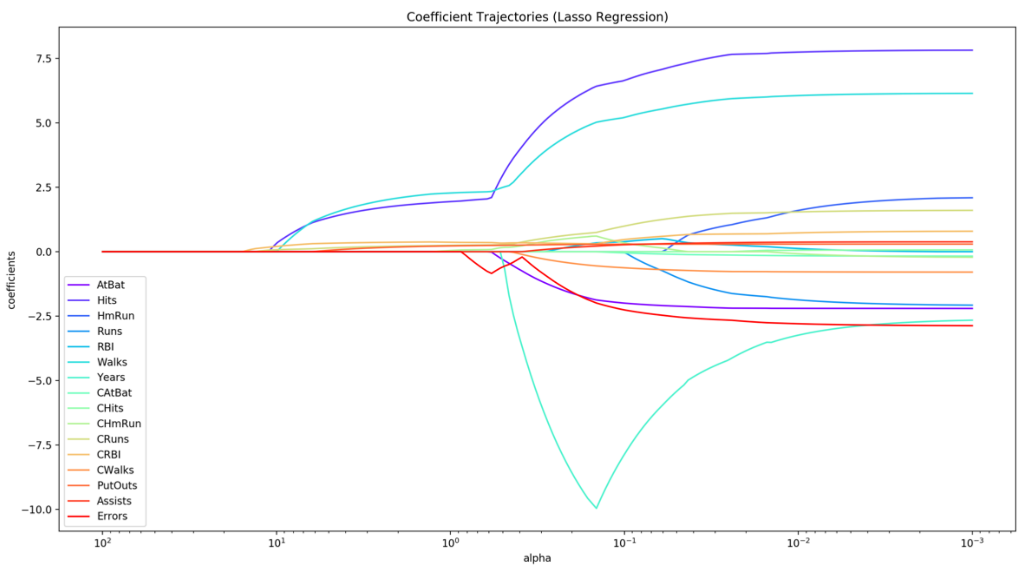
Kevin Kyunggeun Jang
ANLY-590
Homework 0
09/19/18

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.

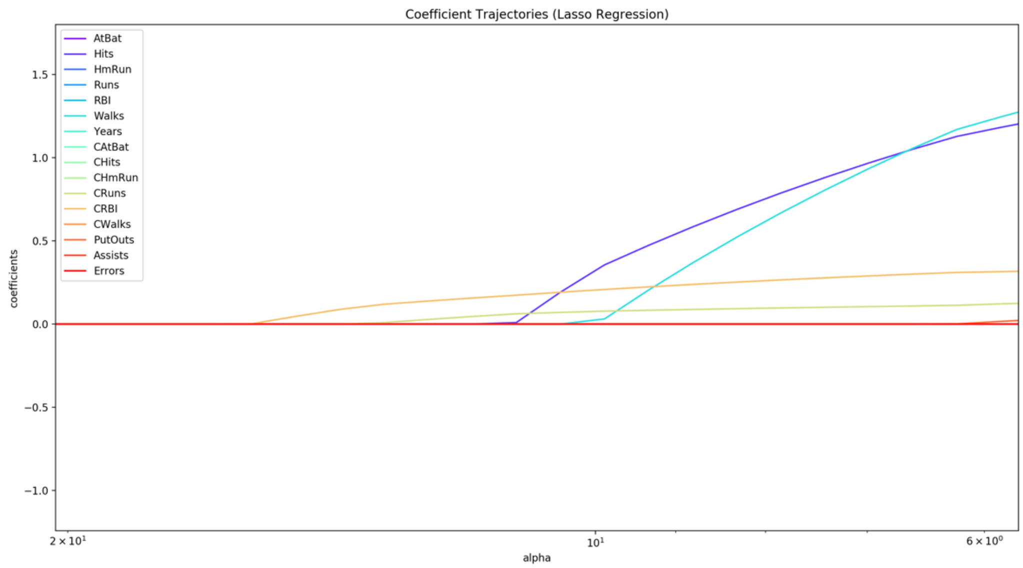
1.1 Use LASSO regression to predict Salary from the other numeric predictors (you should omit the categorical predictors).

- 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: {'$\boldsymbol{CRBI}$', '$\boldsymbol{CRuns}$', '$\boldsymbol{Hits}$'}

- 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: $\mathbf{0.0635481759985745}$

- 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:

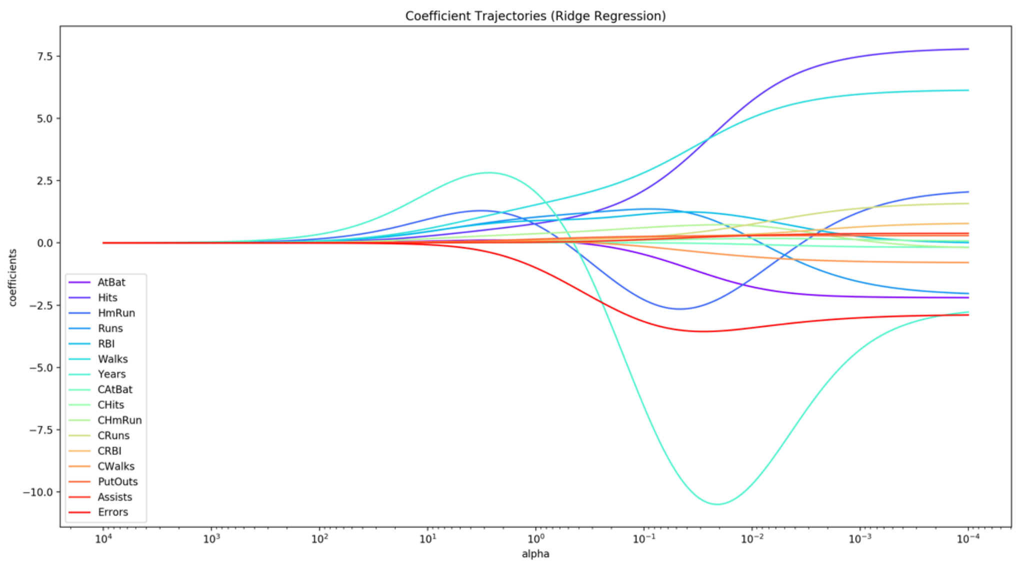
```
array([-1.38848835,  7.76092133,  6.64732999, -1.87452439, -2.85735254,
        4.82400254,  0.        , -0.15602497,  0.        , -2.11727316,
        1.19111661,  1.25130813, -0.45490904,  0.15003829,  0.01929291,
       -4.73926396])
```

(**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 coefficient 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: $\mathbf{0.9547716114208056}$

2. **Short Answer.**

- **Explain in your own words the bias-variance tradeoff.**
  - In my words, bias-variance tradeoff is the phenomena where the quantity of errors occurring by bias and variance, changes in an opposite direction, in other words, changes inverse proportionally. As an example, the case when reduction in bias errors results increase of variance errors.

- Additionally, bias is the type of an error resulted by not considering all the values in the dataset and causes underfitting. Where variance is the type of an error resulted by the models that work too well to catch all the errors from the dataset and causes overfitting.

- **What role does regularization play in this tradeoff?**
    - Regularization plays a role in the bias-variance tradeoff, in a way that it reduces variance which also reduces the chance of overfitting from happening on our models. This can be done by reducing the weights of our predicators or ignoring some of the predicators. Through this, the performance of our model will be more regularized, meaning the variance will be reduced and overfitting will be solved.

- **Make reference to your findings in number (1) to describe models of high/low bias and variance.**
    - In number (1), we used Lasso and Ridge regressions to achieve the effect of regularization. Through testing with different alpha values, we tried to penalize the weights of our numeric predicators and monitor how weights are changing over different alpha values. As a result, on Lasso with the optimal alpha value returned by 'LassoCV', reduced the size of predicators from sixteen to fourteen (two predicators resulted zero coefficient).
    - ```
      lasso optimal MSE = 130984.09084628598
      lasso train MSE = 79952.41965228014
      ```
    - ```
      ridge optimal MSE = 152734.08842667344
      ridge train MSE = 77219.72780718851
      ```
    - By looking at the MSE (Mean Squared Error) values above, we can notice that our model was overfitting, meaning low bias and high variance. However, through regularization techniques, the tradeoff happened by reducing the variance.