

Linear Regression

Project Write-up

Fri., April 2, 2021

Abstract:

Like most things in life, professional golf is increasingly being analyzed through data and statistics. This is evident during weekly telecasts. This project is to see how closely I can come to predicting annual earnings through the use of linear regression and published statistics. Examples of these statistics are driving distance and average putts per hole.

Data Description:

The data used was scraped using Beautiful Soup from [ESPN.com](https://www.espn.com). In all 22 fields were gathered in this manner. Years included were 2010 through 2020. 2016 data was not available and not used as a result.

Algorithms:

Linear regression algorithms were used for simple, LASSO, Ridge, and polynomial regression.

Tools:

Beautiful Soup for Web Scraping
All data manipulation was completed in a Jupyter notebook and Pandas dataframe
Matplotlib and Seaborn for plotting
Additional Python libraries regression model calculations.

Communication:

I will be presenting results to the Regression cohort on Friday, April 16, 2021. A PDF of the presentation is also included on my personal GitHub repository.