HM2

Writeup Xiang Li

Data

As someone entirely unfamiliar with American football and unaware of what the Super Bowl entails, finding suitable data posed my initial challenge. I attempted to download data from Pro Football, but the abbreviations within the data were indecipherable to me. I also tried crawling data from NFL.com using a web scraper, but that too ended in failure. In my moment of despair, I am grateful to my fellow classmates who pointed me towards Kaggle, where I found the appropriate data. I obtained the complete names for each data type, facilitating a better understanding of each category. Ultimately, I downloaded the "NFL Team Data 2003-2023" dataset from Kaggle (https://www.kaggle.com/datasets/nickcantalupa/nfl-team-data-2003-2023/data). This dataset encompasses all teams participating from 2003 to 2023, containing 673 rows and 35 columns, including information such as year, team names, number of wins and losses, points, points against, point differential, and more. The dataset is thorough and well-suited for predicting champions due to its ample information. The data is already quite tidy, requiring minimal cleaning. In the end, I opted to compare the Lasso and Ridge models initially, selecting the more accurate one for predicting the final outcome.

Exploratory Analysis

I conducted an analysis of the Lasso and Ridge models, obtaining their Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 Score. For MSE and MAE, lower values are better, while a closer to 1 R^2 Score is preferable.

Results showed as below:

Model	MSE	MAE	R^2 Score
Lasso	0.0430	<u>0.1159</u>	0.9959
Ridge	<u>0.0424</u>	0.1164	<u>0.9960</u>

It is evident that Ridge secured a marginal victory.Therefore, I have chosen to proceed with the Ridge model for further analysis.

Results

Based on the model's predictions, the team **Kansas City Chiefs** is foretasted to win the Super Bowl championship.