Dr. Tony Diana
DATA 602 Introduction to Machine Learning
Practice Exercises I Lecture Week 5

Lasso

- Import the libraries including LASSO and Ridge
- Load the 'mlb.csv' dataset
- Use salary as the target
- Drop 'League', 'Division', and 'NewLeage' from the features
- Split the dataset into a training and a test dataset including 25% of the data
- Fit a LASSO model. Increase number of iterations if there is no convergence
- Provide the coefficients and the intercept of the LASSO model
- Use different values for alpha = (0, 1000, 100)
- Plot the alphas and coefficients with a log X scale using the gca() method
- Provide the top five predictions based on training and test datasets
- Provide the mean squared error and R^2 based on test dataset
- Use the LassoCV algorithm to find the optimal alpha value using cross-validation = 10 with max_iter=100000
- Provide the alpha value of the LassoCV model (tuned model)
- Provide the prediction and mean squared error of the tuned model
- What are the tuned model coefficients? What variables were not used in the model and why?