



Homework 1: Razorback Football Win Margin Prediction

Goal: Create the best regression model to predict the win margin of razorback football games. You will then use your model to predict the win margin for select games during this fall season and will compete with your fellow classmates to see who has the best model. To do this you will implement popular regression machine learning models in python and use performance metrics to quantify the results of the models on testing data. Also, you will investigate the differences between parametric and non-parametric machine learning regression models.

Data: You will be provided data prepared from past Razorback football games ranging from 2010-2024 acquired from [here](#).

Model Inputs: There is a total of 96 provided possible input features such as Arkansas and the Opponents First downs, Completed Passes, Touchbacks, if it's a home game, etc. There are several features provided but not all have to be used in your model. It is up to you to decide which are the most relevant.

Model Outputs: There are two columns at the end of the csv file that provide the game scores for Arkansas and the Opponent. You will need to subtract the Opponent Score from Arkansas's Score to get your desired output (win margin). Your output value will be positive if Arkansas wins and Negative if they lose.

Assignment:

1. Construct and train at least 4 machine learning regression models for predicting the win margin:
 - a. Linear Regression (parametric model)
 - b. Multi-layer perceptron (MLP) – Circumvent the effects of overfitting using k-fold cross-validation (e.g., using 100 foldings)
 - c. Gaussian Process Regression (GPR) – The choice of kernel is very important. You want it to represent how the output will look like. One popular kernel is RBF.
 - d. Random Forest Regression
2. Performance Metrics
 - a. Report training curves (training/validation loss vs. epoch)
 - b. Training time for the models (time/epoch, time til best model)
 - c. Using your test data, plot the predicted value vs time and true value vs time in a scatter plot
 - d. Plot a bar graph of the MSE (or any common regression metric) comparing the four different models
3. Competition
 - a. After the next football game use your best model to predict the win margin based on the game statistics. We will see who has the closest prediction.