Michael Buchar
MAIS 202

Bundesliga Predictor

Deliverable description

1. Dataset

I have chosen the Bundesliga dataset from https://datahub.io/sports-data/german-bundesliga. I would like to implement an iOS app that will get as input an upcoming Bundesliga game and then it will predict the probabilities of each team winning as well as the probability of a draw. I will use 5 datasets from the past 5 years and give the more recent ones an exponentially higher weight since recent games are more important than the ones played 5 years ago (or I will let the machine figure out the best correlation). This dataset is suitable since it has a lot of parameters including half-time scores, total shots, total shots on target by each team etc…

2. Methodology

This code will be developed in Python, a language widely used by the machine learning community. There are around 20 features from which the algorithm will learn and try to predict the best outcome. Since this is a multiclass problem, I have decided to use XGBoost, which is widely used by the Kaggle community and performs extremely well in their yearly March Madness competitions (https://www.kaggle.com/c/mens-machine-learning-competition-2018). XGBoost is a software library which provides a gradient boosting framework and has been recommended by many machine learning professionals.

The program will take as input two parameters: the home team and away team and will output the probabilities of “H” (home team win), “A” (away team win), “D” (draw).

I am planning on an extension which would take two more parameters (the half-time home team goals and the half-time away team goals) and again try to predict the winner of the game. To conclude, I would like to implement a simple iOS app with a nice design, where one can input two Bundesliga teams and receive the output mentioned above.