Machine Learning Engineer Nanodegree

Capstone Project Proposal

Pothuru Yaswanth

February 2nd 2019

Proposal:

Chess (King-Rook vs. King)

Domain Background:

History:

Rook endgames are the most common type of endgames there is in the game of chess. These endgames take place in about in 50% of all the games. The main idea is checkmate of black king with white king and white rook. This should be done in less than 16 steps.

Chess endgames are complex domains which are enumerable. The game theoretic values stored denote whether or not positions are won for either side, or include also depth of wins (number of moves).

The relevant paper related to this M. Bain. "Learning optimal chess strategies", ILP 92: ICOT TM-1182, S. Muggleton, Institute for New Generation Computer Technology, Tokyo, Japan. [Web Link]

Problem Statement:

The main aim of my project is to predict the outcome of the king-rook vs king endgame. For this I selected the dataset from UCI (https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29)
So my goal is to predict the outcome of the game whether the white king will win the game in less than 16 moves or not. Here I am using classification models to find the accuracy of each model and select the model which will have high accuracy. Here the input parameter is given in the form of training data.

Datasets and input parameters:

Here the dataset will have categorical values and integer values. There will be total of 28056 instances, and the total number of attributes is 6. The dataset here is a multivariate.

Attribute information:

White king file (column)
White king rank (row)
White rook file
White rook rank
Black king file
Black king rank
Optimal depth-of-win for white in 0 to 16 moves otherwise drawn {draw, 0, 1,2,....,16}

Solution Statement:

Here I am trying to predict the outcome from the selected data. For doing so we want to use different classification models. Then we will find the accuracy score for each classification model. I explore the dataset with read_cv, and matplotlib.pyplot libraries in this project. By using visualization helps me to better understand the solution.

Benchmark model:

This step will be important because compare your final model with some of them and see if it got better, same or worse. Here Accuracy score will be compare between the models and select the best one.

Evaluation Metrics:

Here I will use accuracy score as evaluation metric. I will be predicting the accuracy score for the selected model. The model with high accuracy score will be the best model out of the chosen models.

Project Design:

The project is composed of the following steps:

Pre-processing:

First task is to read the dataset and perform visualization on it to get some insights about the data. After reading the data clean the data that is removing unwanted data or replacing null values with some constant values or removing duplicates

After data exploration I will split the data into training set and testing set. Then applying the classification models and predicting the accuracy score to the selecting models.

Training and Testing the data:

I want to apply classification models of my own and use them on the data. I want to apply svm, logistic regression, and random forest.

Then I will find the accuracy score for the above mentioned models. For this I will first train the algorithms with the training data, and then carry on testing with the testing data that I split before

Finally, I will declare the model which has the highest accuracy score out of all the chosen algorithms and declare it as the best one.