Machine Learning - Project Proposal

---

1. Doman Background

Chess is a game with countless end game scenarios. This project is derived from the scenario where one player has solely a king and rook for pieces while the other player has its king. This project shares resemblances with that of J.R Quinlans (1983) use of ID3 to classify the King Rook against King Knight, formally known as the (KRKN).

2. Problem

Given the three chess pieces and their coordinates on a chess board, my system will predict whether or not the White player (The one with Rook and King) will win the match in 16 moves or under, given that both players would play optimally. The match is declared a draw if the Black King can capture the White Rook, cannot move on its turn as all possible moves would lead to its capture, or threefold repetition where the same position occurs three times with the same player to move. Note: we leave out the 50 turn rule as this dataset goes to a max of 16 moves.

3. Input and Dataset

The input vector will have 6 attributes, the first two corresponding to the column and row of the White King respectively, while the next two correspond to the White Rook and then finally, the last two to the Black King. The dataset used and its description can be found at this URL : https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29

4. Solution Statement

Due to the inherent nature of a chess board being grid like, Knn Classification will be used on the data set to classify the test set into the 17 different classes. The distance metric for both euclidean and manhattan distance will be used with the expectation that manhattan distance will outperform euclidian due to dimensionality and how a rook moves in 1 dimension at a time.

5. Baseline Model

For the baseline Model, Naive Bayes will be used as this is the traditional approach when no pre-existing historical model can be found.

6. Evaluation Metrics

The evaluation metric will be as an accuracy score, confusion matrix and an algorithm comparison of the three algorithms in box plots format. This is done so we can see an in-depth view of how each algorithm performed.

7. Project design

The process will begin with a stratified k-fold cross validation to ensure the model has sufficient information on all 17 classes. From there we will run Knn Classifier with both distance metrics along with naive bayes to gather the results to show the accuracy score, confusion matrix and finally make an algorithm comparison and decide which performed best.

8. bibliography

J. R. Quinlan. Learning Efficient Classification Procedures and their Application to Chess End Games.Machine Learning: An Artificial Intelligence Approach. 464--482. R. Michalski and J. Carbonnel and T. Mitchell, eds. Tioga, 1983. Palo Alto, CA.

Cafe Scientifique. Distances in Classification. PDF File. July 1st, 2016. http://www.ieee.ma/uaesb/pdf/distances-in-classification.pdf

Bain, Micheal, and Arthur van Hoff. “Chess (King-Rook Vs King) Data set”. UCI Machine Learning Repository.https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29