Machine learning and Data analysis

CW1 Report

Introduction

The following report will discuss the implementation of a supervised machine learning algorithm, providing a thorough summary of the processing of data as well as the implementation of the machine learning model, while also reporting on the accuracy of the implemented model. Information on the chosen dataset and steps required to sanitise this data to be used with a supervised machine learning model will be discussed while also going into detail the steps required to implement a support vector machine (SVM) model, a Neural network model or a deep learning model. Finally, the results of the chosen model will be reported, and any steps taken to tune or improve the results of the model will also be reported on with a short discussion on the importance of the data processing used in this report, as well as discussing the chosen machine learning model and what made it effective implemented with the chosen dataset.

Description of data and the problem

The chosen dataset relates to information collected from players of the popular video game league of legends, which is a multiplayer online battle arena (MOBA) videogame where two teams of five fight to destroy each other’s base. Players eliminate other players and obtain objectives to earn currency which can be used to gain an advantage over other players, with the eventual goal of the game to destroy the other teams base resulting in a victory. The data collected is from the first 10 minutes of roughly 10 thousand games, with each instance of the dataset relating to a game of league of legends. Since the average length of a game of league of legends is usually 30-40 minutes long, the first 10 minutes of data refers specifically to the “early game” as it is referred to. This “early game” is arguably one of the most impactful parts of a game of league of legends and the choices made during this 10-minute window could decide the outcome of a game.

The chosen dataset has roughly 10 thousand instances of data with each relating to the first 10 minutes of an individual game of league of legends, each with 40 different attributes, with numerical data on gold earned, experience earned, how many objectives have been taken by a team and many other important pieces of information. Minimizing the dataset to information on a game from the first 10 minutes helps to keep the size of the dataset small while still being able to include thousands of games worth of data.

Therefore, it makes sense that the information held in the data structure could be used to predict whether or not a team may or may not win a game of league of legends. Most importantly the dataset specifically already contains a variable which has a record of whether or not a team won or lost their game, and can be used as a target value for a supervised machine learning algorithm to predict the outcome of a game using the information in the dataset.

The problem looking to be solved is can the outcome of a game of league of legends be predicted using machine learning using data collected within the first 10 minutes of a game. Given all instances of the dataset are numerical, use of a support vector machine (SVM) model would be appropriate.