Predicting the English Premier League Table using a Neural Network

Date: 1/2/2020

Version 0.0.1

Author: Kunal Buty

# 1: Data Preparation

## 1.1: Training Data

* Training\_Data will have the following shape: [[Features],[Labels]]
  + The following is an example of the Training\_Data Array
  + [

[[80,81,85,73,77,74],[80,80,84,81,83,79],[63,69,70,75,75,70]],

[1,0,2]

]

* The following Features will be used to train the Neural Network:
  + Home Team Attack Rating
  + Home Team Defense Rating
  + Home Team Midfield Rating
  + Away Team Attack Rating
  + Away Team Defense Rating
  + Away Team Midfield Rating
* Labels will be one of the following:
  + 0 (Draw)
  + 1 (Home Win)
  + 2 (Away Win)

## 1.2: MySQL Tables

* Training\_Set\_0.0.1
  + This will contain 5 seasons worth of results (1900 Matches)
  + Team Ratings will be determined by taking team’s FIFA stats (DEF, MID, ATT) from the FIFA version corresponding to the year of the match

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Home\_Team\_DEF | Home\_Team\_MID | Home\_Team\_ATT | Away\_Team\_DEF | Away\_Team\_MID | Away\_Team\_ATT | Result |
| 80 | 81 | 85 | 73 | 77 | 74 | 1 |
| 80 | 80 | 84 | 81 | 83 | 79 | 0 |
| 63 | 69 | 70 | 75 | 75 | 70 | 2 |

## 1.3: Match Data

* The following Data will be required in order to create a training/test set:
  + 2010-2015 Match Results
  + FIFA 10-FIFA 15 Team Ratings

# 2: Neural network Parameters

## 2.1: Input Layer

* Since we have 6 features, our input layer will have 6 neurons.

## 2.2: Output Layer

* Since there are 3 possible outcomes (0, 1, 2), our neural network must have 3 neurons.

## 2.3: Hidden Layers

* Since I have no idea how to do this, I will find the accuracy for each combination of number of hidden layers and number of neurons in each hidden layer.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 Hidden Layer | 2 Hidden Layers | 3 Hidden Layers |
| 4 Neurons |  |  |  |
| 5 Neurons |  |  |  |

## 2.4: Activation Functions

* I will use softmax on the output layer. This will choose the highest value of the 3 output neurons, which will give me the option with the most probability
* For hidden/Input layer, I will use ReLU. I don’t know why, but I will play with this and try other ones too.