Week 7:

*What I've done this previous week*:

- Researched a similar project which models premier league teams and makes predictions. Understood from this that I need to add an output to the data in order to supervise the network’s learning.

- Took time to study the inputs that I plan on feeding to my network and realised that a small number of them contained information that would not be available prior to a match. I would therefore not be available to pass these as input if the network was making a prediction on a future match. These inputs were removed and replaced by other relevant information that the network could use to find underlying dependencies in the data.

- Rewritten the following classes:

- CreateTables – in order to read all the data into the training data table. The purpose for this is that all the data can be normalised on the same scale, and then separated into the training and test data tables once this is complete.

- CleanDatabase – to remove the columns in the database tables that aren’t needed to train or test the network. This also included creating numerous helper methods for altering the types of values in the database tables as the data from the CSV files contained incorrect values and had to be read in as type VARCHAR.

- Rewrote the entire Normalise class – now successfully normalises all numeric columns (with the exception of the player’s height).

- Updated StandardiseNonNumeric class – previously contained individual methods to standardise each non-numeric column, rewrote to one dynamic method.

- Added checks for bad parameters to the majority of methods.

- Changed SQL Statements to PreparedStatements.

- Created ReadFiles class which reads in the CSV files to the database (for previous weeks it was loaded manually).

- Created FinaliseDatabase class which performs following functions: Adds a primary key “match\_id” to uniquely identify values, adds data from 2017-18 to the test database, and creates two further tables for the winners and losers of each match which contain an output column with values 1 for the winner and 0 for the loser of each match (these will be linked by the match\_id).

- Reorganised the project structure into well-defined packages.

- Debugged all database code so far.

- Planned the order of execution of the program which now successfully reads in the files, filters the information needed, normalises numeric data (min-max), standardises non-numeric data with random float values between -1 and 1 and splits the data into relevant database tables.

*What I will do this week*:

- One final issue with standardising the numeric age of each player. I will debug this and the database side of the project will be complete.

- Revisit the neural network classes and tailor them to fit this project now I possess a greater understanding.

- Begin the process of training the network.

*What we should discuss*:

- I have prepared a list of questions which I will write up in next week’s log as what we discussed this week.

*Previous meeting summary*:

- Different options for organising data to feed it to the network.

- Progress so far and what was necessary to complete project on time.