

Deep Learning Intermediary Report: Regression on Soccer Player's Skills

Arturo Perez

The objective of this project will be to predict the “overall rating” of a Soccer player given attributes about his skills using the European Soccer Database. The problem boils down to be a regression problem as I want the model to output a number in the range [1,100) based on 26 skill attributes for each player. The database provides me with the actual Overall Rating of each player which I will use as my labels.

An example of the problem:

FIFA Player ID: 179813

Crossing: 70

Finishing: 85

Heading Accuracy: 86

Dribbling: 81

Short Passing: 77

...

...

...

Overall Rating: 85

What I am thinking so far:

Since predicting the exact Overall Rating for each player is too an idealistic of a situation, I am rather trying to see if my predictions can fall in a range around the actual Overall Rating.

For example:

Range = ± 7	Predicted Rating: 92	Actual Rating: 89	Prediction: Success
Range = ± 7	Predicted Rating: 92	Actual Rating: 80	Prediction: Failure

I will attempt different methods to maximize the accuracy of my predictions (therefore minimizing the range) and these are some ideas I have so far: running the model with all 26 skill attributes, running the model on selected subsets of the skill attributes, running the model on randomly selected subsets of skill attributes. The actual implementation may vary as I move forward in the project and see what works and what does not work.

Progress: So far I have read in the data from the MySQL database using Pandas, and am still in the process of preprocessing and partitioning the data into something I can use. My next move will be to begin implementation of the model most likely using TensorFlow.