Artificial Neural Networks

Hands-On Project: Building an Artificial Neural Network

Due date: Friday, December 16.

Building an artificial neural network to estimate the overall quality of a football player.

This activity aims to follow the construction process of a deep artificial neural network for a classification problem using the Fifa19 dataset*. The goal is to estimate the overall quality of football players based on their skills as a classification task.

This dataset has been downloaded from Kaggle: https://www.kaggle.com/karangadiya/fifa19; license CC BY-NC-SA 4.0: https://creativecommons.org/licenses/by-nc-sa/4.0/, and refers to FIFA 2019 edition videogame players.

The dataset initially contains 18,207 instances with 89 attributes each related to football players. After a preparing and cleaning process, we end up with 16,122 instances with 17 attributes regarding football player skills: Crossing, Heading Accuracy, Short Passing, Volleys, Dribbling, Curve, Free Kick Accuracy, Long Passing, Ball Control, Reactions, Shot Power, Stamina, Long Shots, Aggression, Positioning, Vision, and Composure. These attributes have been chosen because their values correlate greater than 0.4 or less than -0.4 with the overall score (output), which has been quantile-based discretized into four classes or categories: *Poor* football players with overall scores in [46, 62], *Intermediate* for values in [63, 66], *Good* players in [67, 71], and *Excellent* players for overall values in [72, 94].

Students, in **groups of three people**, will perform the following tasks:

1. **Prepare the environment for Python 3 with Tensorflow 2 and Keras.** You can choose any development environment, such as PyCharm® IDE (https://www.jetbrains.com/pycharm/), Atom® (https://atom.io), Anaconda® distribution (https://www.anaconda.com/download/), although Google Colab® is encouraged since it requires no setup: https://colab.research.google.com/.
2. **Clean and prepare the dataset.**
   - Download the raw-data file *FootballPlayerRawDataset.csv* and the notebook *PreparingFootballPlayerDataset.ipynb* from the Moodle platform, section *Practical Assignment*, Hands-on project Units 1 and 2: Building an artificial neural network.
   - Upload the raw-data file and the notebook under your *Colab Notebooks* folder in your GDrive.
   - Open the notebook and configure the file path INPUT_FILE_NAME to point to the raw-data file location, and ATT_FILE_NAME and ONE_HOT_ENCODED_CLASSES_FILE_NAME to point to the desired folder of the resulting files, i.e., the file with the attributes (inputs) and the file with the one-hot-encoded-classes, respectively.
   - Execute the notebook to clean and prepare the data and obtain the resulting files to feed the neural models. Pay attention to the actions performed to better understand the dataset. The result of this process should be two CSV files: FootballPlayerPreparedCleanAttributes.csv and FootballPlayerOneHotEncodedClasses.csv. The former contains prepared and clean instances for the attributes (predictors). The latter includes the corresponding one-hot encoded classes

for the overall quality of football players, grouped into four categories: poor, intermediate, good, and excellent.

3. **Construct a deep neural network**. Write a notebook that first implements the data loading process of the two .csv files, Attributes, and Classes. Then split the dataset into three partitions: 80% of the whole dataset for training, 10% for development testing, and the remaining 10% for final testing purposes. Finally, follow the deep-neural-network construction process to find the neural architecture and other hyperparameters that achieve the best performance in classification accuracy. Consider a Bayesian error of 10% (minimum error); i.e., the human error that soccer scouts make when predicting the quality of football players from their skills. You can use the notebook that implements the deep neural model in Keras for the median house value studied in class as a starting point for this task.
4. **Optionally**, it is possible to modify the data cleaning and preparation process (FootballPlayerPreparedCleanAttributes.csv) to check whether you can find a neural model that improves the results previously achieved. If successful, this work will be considered in the grade of this assignment. It is mandatory to keep the same partition for the classes to predict and the problem of classifying football players into the four above-mentioned classes.
5. **Write a report**, in Spanish or English, describing the actions performed during this activity and the final results. The notebook developed in the previous task may be helpful to this end. The structure of this report is described below. The correctness of the construction process followed is essential. It is also necessary to adequately employ the training, development, and final testing datasets at the right time.
6. **Send the report as a single pdf file**, via Moodle, no later than December 16. Make a single upload for all group members and keep the source code (notebook) in case the instructors require it during the revision process.

The structure of the report to write is the following:

1. **Cover page.** Include a cover page with title, authors, email, course, and date.
2. **Introduction.** Explain the problem to solve and the datasets.
3. **Design process**. Describe the process you followed to reach the final results, showing the intermediate network architectures used and the rest of the hyperparameters employed. Explain your design decisions, justifying why you tested each new neural model. Show the performance (accuracy) of each intermediate model. If applicable, point out when and how you made changes in the data cleaning and preparation process.
4. **Final results.** Describe the ultimate neural network solution, clearly showing all the hyperparameters used. Display how the accuracy changes during the training process of this model and the accuracy achieved for the final test set.
5. **Conclusions**. Summarize your work and the most relevant results.

NOTE: Other group sizes (two or four people) may be exceptionally accepted. Ask for permission by email to martin.molina@upm.es and daniel.manrique@upm.es indicating the reasons.

* The details of this dataset are described in https://www.kaggle.com/karangadiya/fifa19.