Deep Learning

Practical Project: Building a Deep Neural Network

Due date: April 15, 2020.

Building a deep artificial neural network to estimate the approximate location of housing blocks.

The objective of this activity is to follow the construction process of a deep artificial neural network for a classification problem using the California Housing Prices dataset.

This classification problem consists of estimating the approximate location of housing blocks. The approximate location is represented with a discrete variable called ocean_proximity which may have one of four possible values: NEAR BAY, <1H OCEAN, INLAND, and NEAR OCEAN.

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

1. **Prepare the environment for Python 3 with Tensorflow 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. **Download the attributes and class files.** Download the files OceanProximityPreparedCleanAttributes.csv and OceanProximityOneHotEncodedClasses.csv from the Moodle platform, section Practical Project submission located at the very bottom of Unit 1. The attributes file contains the clean and prepared data for the nine attributes (predictors) ready to feed a neural network. It consists of 20,428 rows (plus the header) and nine columns labeled as longitude, latitude, median age, total rooms, total bedrooms, population, households, median income, and median house value. The class file contains the one-hot encoded values for the class variable ocean_proximity. The following are some of the actions performed on the original public dataset HousingRawDataset.csv to prepare the attributes and classes files:

    - Remove the instances corresponding to the value ocean_proximity=ISLAND because there is not enough data for the learning process (only five instances).
    - Remove instances with missing values.
    - Randomize the order of appearance of rows.
    - Max-min scale the data.
    - Encode the ocean_proximity values according to a classification task.

3. **Construct a deep neural network**. Write a notebook implementing first 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 out the neural architecture and its hyperparameters to achieve the best performance regarding the

accuracy in classification. You can use the notebook implementing the deep neural model with ReLU in Keras for the median house value studied in class as a starting point for this task.

4. **Write a report** 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 important to adequately employ the training, development testing, and final testing datasets at the right time.
5. **Send the report as a single file in pdf**, via Moodle, no later than April 15. Maximum allowed size: 50 MB.

The structure of the report to write is the following:

1. **Cover page.** Include a cover page with a title, authors, email, course, and date.
2. **Introduction.** Explain the problem to solve and the datasets.
3. **Design process**. Describe the process that 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 tried each new neural model. Show the performance (accuracy) of each intermediate model.
4. **Final results.** Describe the ultimate neural network solution, showing clearly all the hyperparameters used. Plot or write 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.

NOTES:

- Only one of the workgroup members must perform the submission.
- Other group sizes (one or two people) may be exceptionally accepted. Ask for permission by email to martin.molina@upm.es and dmanrique@fi.upm.es indicating the reasons.