Artificial Neural Networks

Hands-On Project: Building an Artificial Neural Network

Due date: Friday, December 15.

Building an artificial neural network to predict whether the customer will cancel the hotel reservation.

This activity aims to follow the construction process of a deep artificial neural network for a classification problem using a hotel reservation dataset*. The goal is to predict whether a customer will honor the hotel reservation or cancel it based on his profile (e.g., repeated guest, number of previous cancelations) and the features of the reservation (e.g., number of children, car parking space).

The raw dataset contains 36,275 instances and 19 variables. Each observation is referred to as a hotel booking due to arrive between 2017 and 2018. There are bookings that effectively arrived and bookings that were canceled.

After preparing and cleaning the dataset, we end up with 36,270 instances and 16 attributes each: no. adults, no. children, no. weekend nights (Saturday or Sunday), no. week nights (Monday to Friday), type of meal plan, required car parking space, room type reserved, lead time (number of days between the date of booking and the arrival date), arrival month, arrival date (date of the month), market segment type (online, offline, other), repeated guest, no. previous cancellations, no. previous bookings not canceled, avg. price per room (avg. price per day of the reservation in euros), and no. special requests (e.g., view from the room). The *booking status* (the variable to predict) can take two values (classes or categories): 0 (canceled) and 1 (not canceled).

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

1. Prepare the environment for Python 3 with Tensorflow 2 and Keras. You can choose any development environment, although Google Colab® is encouraged.
2. Clean and prepare the dataset.
   - Download the raw-data file *HotelReservationsRawDataset.csv* and the notebook *PreparingHotelReservationsDataset.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 paths INPUT_FILE_NAME, ATT_FILE_NAME, and OUTPUT_FILE_NAME to point to the raw-data file location and to the resulting files with the attributes (inputs) and the target outputs, 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: HotelReservationsPreparedCleanAttributes.csv and HotelReservationsOutput.csv. The former contains the prepared and clean instances with the attributes (predictors). The latter includes the corresponding target classes, labeled as 0 (canceled) and 1 (not canceled).
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% 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 5% (minimum error); i.e., the human error that hotel managers (experts) make when predicting if customers will cancel their bookings. You can use the notebook that implements the deep neural model in Keras to estimate the median house value studied in class as a starting point for this task. Note that we are dealing with a binary classification problem, therefore, we only need one output neuron with the Sigmoid activation function. Moreover, the loss function should be the *binary* crossentropy and the *binary* accuracy as a metric.

(https://keras.io/losses, https://keras.io/metrics)

4. Write a report, in Spanish or English, describing the actions performed during this activity and the results. The notebook developed in the previous task may be helpful in this regard. 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.
5. **Send the report as a single pdf file**, via Moodle, no later than December 15. 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 get the final results, showing the intermediate network architectures and the rest of the hyperparameters used. Explain your design decisions, justifying why you tested each neural model. Show the performance of each intermediate model.
4. Final results. Describe the ultimate neural network, clearly showing all the hyperparameters used. Display how the accuracy changes during the training process of this model. Calculate and analyze the confusion matrix for the final test set.
5. Conclusions. Summarize your work and the most relevant results.

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

* N. Antonio, A. Almeida, L. Nunes. (2019). Hotel booking demand datasets. Data in Brief, 22: 41-49.