Artificial Neural Networks

Practical Project: Building an Artificial Neural Network

Due date: Sunday, December 13.

Building an artificial neural network to estimate the quality of a wine.

This activity aims to follow the construction process of a deep artificial neural network for a classification problem using the Wine Quality dataset*. The goal is to estimate wine quality based on physicochemical tests as a classification task.

This Wine Quality dataset has been downloaded from the UCI repository https://archive.ics.uci.edu/ml/datasets/wine+quality. This dataset refers to the white variant of the Portuguese "Vinho Verde", wine from the north of Portugal.

The dataset initially contains 4,898 instances with 11 attributes, each related to objective tests: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, and alcohol. The label quality (output) has 11 classes based on sensory data (median of at least three wine experts' evaluations). Each expert graded the wine quality between 0 (horrible) and 10 (excellent). However, the data is unbalanced. There are many more normal or intermediate wines than excellent or poor ones. Therefore, it is a good idea to group wines into less than 11 quality classes. Asi que se reducen las clases de vinos a 3: buenos, malos y medios.

(del 0 al 10) Pero los datos están muy desbalanceados,

Students, in **groups of two 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 *WineQualityRawDataset.csv* and the notebook *PreparingWineQualityDataSet.ipynb* from the Moodle platform, section *Practical Assignment*, Practical project: building an artificial neural network. CON ESE NOTEBOOK Y ESE FICHERO YA SE GENERAN LOS ATRIBUTOS LIMPIOS Y PREPARADOS, CON LAS CLASES YA CODIFICADAS COMO ONE-HOT ENCODING
   - 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.

     Este apunta al conjunto de datos en bruto Wine....cvs

     Y estos dos ficheros serán las entradas y salidas de la red de neuronas:
   - 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 understand the dataset better.

El proceso anterior nos proporcionará estos 2 ficheros al ejecutarlo:

11 ENTRADAS

The result of this process should be two CSV files: WineQualityPreparedCleanAttributes.csv and WineQualityOneHotEncodedClasses.csv. The former contains prepared and clean instances for

CONTIENE LAS 3 CLASES
(Necesitaremos 3 vectores de salida)

the eleven attributes (predictors). The latter includes the corresponding one-hot encoded classes for the wine quality, grouped into three categories: poor, regular, and excellent. You can modify or adapt the notebook provided if you think you can improve the proposed data cleaning and preparation process.

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 other hyperparameters that achieve the best performance in classification accuracy. You can use the notebook implementing the deep neural model 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 necessary 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 December 13.

Datos repartidos:

(EL NOTEBOOK ÚLTIMO EXPLICADO)

Memoria en PDF que explique todo esto

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. Display how the accuracy changes during this model's training process, and the accuracy achieved for the final test set.
5. **Conclusions**. Summarize your work and the most relevant results.

Describir el proceso para alcanzar el resultado mostrando las redes intermedias usadas y los hiperparámetros.

Explicar las decisiones, justificar por qué habéis intentado para modelo neuronal.

Mostrar el accuracy de cada modelo intermedio

Finalmente, mostrar el accuracy del final test set del modelo neuronal que creamos que funciona razonablemente bien

NOTE: Other group sizes (one or three 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 P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. (2009) Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4): 547-553.