Learning with Artificial Neural Networks

Practical Work 03 – Supervised Learning and Speaker recognition using Neural Networks

Professor: Andres Perez-Uribe
Assistants: Benoît Hohl, Shabnam Ataee
Emails : prenom.nom@heig-vd.ch

Goals:
- Learning the methodology of model selection.
- Understanding the problem of overfitting and the criteria that help us avoid it.
- Having hands-on experience with Multilayer Perceptrons (MLP).

1. Introduction

During the previous practical work sessions, we provided you with a series of notebooks to explore the workings of an artificial neuron (e.g., the Perceptron model), neural networks (Multi-Layer Perceptrons or MLPs) and Backpropagation, which allows those system to "learn from examples".

During this practical work you will apply the general data-driven modelling methodology to develop algorithms capable of speaker recognition from recorded sounds. You will use the cross-validation approach to evaluate the performance of the trained neural with the aim of selecting a final model to deal with the problem at hand. The selection of a model consists on finding a model of appropriate complexity (e.g., which is related to the number of parameters or synaptic weights) and configuration (e.g., type of activation function, learning rate, momentum rate, training iterations, etc). Finally, you will evaluate the final performance of your models using the appropriate measures (e.g., accuracy, F1-score, confusion matrix)

2. Hold-out validation vs cross-validation

You will find two notebooks that explore the use of hold-out validation and cross-validation in a simple classification problem of synthetic data. Please, explore them and analyze the results you obtained, as discussed in the lessons.

3. Model building

When training a neural network to solve a problem, e.g., to develop a classification system, you will need to evaluate diverse models (neural net configurations, complexities, diverse parameters, etc) and select the "best" one. The **9_model_selection.ipynb** notebook presents a methodology iterating over the number of epochs (learning iterations) and number of hidden neurons (model complexity). When selecting the final model, that is, defining the number of epochs for training and the number of hidden neurons, you will need to evaluate the performance of the final model, by cross-validation, and you might also compute the confusion matrix, which illustrates if the system confuses certain inputs while attempting to classify them.

4. Speaker recognition experiments

You will be provided with a database of vowels spoken by men, women and children (of 3, 5 and 7 years old). The task will be to train artificial neural networks to recognize the speaker having produced the given sounds and evaluate its performance (e.g., by cross-validation). The file vowels.zip contains the sounds in WAV format. They have been collected by the team of Prof. Peter Assmann of the School of Behavioral and Brain Sciences of Texas University in Dallas. You will also find a set of synthetic sounds corresponding to each vowel and each speaker. Please read the file 0_README.txt for more information.

This experiment is the continuation of the practical work on data exploration but this time instead of visualizing data and trying to classify vowels recordings with your own rules or with KNN, you will use a Multilayer Perceptron and you will choose the best model based on a performance assessment of different models' architectures and hyperparameters.

We suggest you use the model_selection notebook to have a skeleton of code for the model selection you have to perform for the problem of speaker recognition. You may also reuse some code of other notebooks.

Report

For each of the following experiments, provide a brief description of the number of observations of each class, the features being used to train the model, the procedure (explain) for selecting the final model (e.g., use the model_selection.ipynb notebook), the description of the final model and its performance evaluation (i.e., provide the cross-validation results, the confusion matrix and the F-score). Comment your results.

1. Man vs Woman. Use only the natural voices of men and women to train a neural network that recognizes the gender of the speaker.

2. Man vs. Woman vs. children (3 classes)

3. Design a final experiment of your choice (e.g., using your own voice or the synthetic voices).