

# Overview of Archived Materials

Telmo Amaral

31th January 2014

Project “Reusable Deep Neural Networks: Applications to Biomedical Data”  
(PDTC/EIA-EIA/119004/2010)

Neural Networks Interest Group

Instituto de Engenharia Biomédica (INEB)  
Rua Dr. Roberto Frias, 4200-465, Porto, Portugal



# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Papers and Reports</b>	<b>3</b>
<b>3</b>	<b>Experiments</b>	<b>3</b>
3.1	Eclipse and PyDev . . . . .	4
3.2	Requirements . . . . .	4
<b>4</b>	<b>Misc</b>	<b>4</b>
	<b>References</b>	<b>5</b>

## 1 Introduction

This document presents an overview of the materials that resulted from my work in the project, stored on the computer that I used in the lab. The materials are stored in the Linux partition of the computer's disk, in the directory shown below. The best way to access it is to boot the computer into Ubuntu 12.04 and log into the *Telmo* (administrative) account.

```
/home/telmo/Documents/rdnn/
```

This directory contains a number of subdirectories, namely *experiments/*, *misc/*, *papers/*, and *reports/*. The contents of these subdirectories are explained in the following subsections.

## 2 Papers and Reports

The *papers/* and *reports/* directories contain the source files for each of the papers that I have submitted to conferences [5, 3, 6] and for a number of technical reports, some of them already sent to the FCT [1, 4, 2], some others waiting to be sent. All of these documents were created using LyX, but if necessary they can be saved as .tex files from within LyX, via File > Export > L<sup>A</sup>T<sub>E</sub>X (pdflatex). The latest output version of each document is also included in PDF.

## 3 Experiments

The *experiments/* directory contains three subdirectories, namely *1\_micai/*, *2\_icpr/* and *3\_ijcnn/*. Each subdirectory corresponds to one of the papers that were submitted and contains the code, prepared data and results associated with the experiments reported in that paper.

Within each subdirectory, there is one or more main Python scripts used to run experiments and to process results. These main scripts are the natural starting point to understand the code. They have names that start with *main* and a file extension *.py*. The *1\_micai/* subdirectory contains one main script per data set: *main\_script\_adult.py*, *main\_script\_cb4x4(10)\_4000.py*, etc. The *2\_icpr/* and *3\_ijcnn/* subdirectories contain only one main script each, named *main.py*.

Within each subdirectory, besides the main script(s), there are three Python files that contain functions written from scratch, namely: *data\_handling.py*, with functions used to handle data (prepare, load, save, etc.); *execution.py*, with functions that control the repeated execution of experiments; and *plotting.py*, with functions used to create plots and perform other tasks that involve processing results.

Within each subdirectory, there are also some files containing 3rd-party code that was modified for our purposes, namely: *deep\_learning.py*, with a modified version of the deep learning code originally written by Hugo Larochelle; and *classification.py* and *generic.py*, modified versions of files originally from the MLPython library.

The most important functions in the *.py* files mentioned above are documented by a “docstring” comment. This comment is located at the top of each function’s definition and explains what the function does and what each of its arguments means. Numerous additional comments are included in the body of the function definitions.

Within each subdirectory, there is a *data/* subdirectory, containing prepared data (that resulted from calling data preparation functions), as well as a *results/* subdirectory, containing files with results output by the experiments and plots created from those results.

### 3.1 Eclipse and PyDev

Eclipse with the PyDev plug-in is a good integrated environment for Python, allowing easy debugging with breakpoints, easy renaming of variables, etc. The experiment directories *1\_micai/*, *2\_icpr/* and *3\_ijcnn/* are actually Eclipse/PyDev projects that can be imported into Eclipse via File > Import > General > Existing Projects into Workspace. Version 3.7.2 of Eclipse was used, on Ubuntu 12.04.

### 3.2 Requirements

The code was developed using Python 2.7.3. The most important Python packages that are used in the code are Numpy (version 1.8.0 used), SciPy (version 0.13.2 used), multiprocessing (version 0.70a1), and Matplotlib (version 1.1.1rc used).

The MLPython library is already included as a subdirectory where necessary. Nevertheless, it requires that an environment variable `MLPYTHON_DATASET_REPO` be created and set to the directory where data sets that are automatically downloaded should be stored. In Eclipse/PyDev, environment variables can be created and set via Window > Preferences > PyDev > Interpreter - Python > Environment > New.

In case OpenBLAS is installed on the system (in Ubuntu 12.04 this means that the *libopenblas-base* and *libopenblas-dev* Ubuntu packages are installed), then an environment variable `OMP_NUM_THREADS` should be created and set to 1, otherwise multiprocessing will not work properly. OpenBLAS will probably be installed if you’re using the Theano library for other code development tasks.

## 4 Misc

The *misc/* directory contains two subdirectories: *raw\_data/* and *3rd\_party\_code/*.

The *raw\_data/* subdirectory contains original data that is read by some of the data preparation functions (that is, functions defined in the *data\_handling.py*

files). So, these original data may be needed, if you decide to re-run data preparation functions.

The *3rd\_party\_code/* subdirectory simply contains a backup of 3rd-party code that is used in the experiments, such as the deep learning code developed by Hugo Larochelle and the MLPython library.

## References

- [1] T. Amaral. Experiments with a restricted Boltzmann machine. Technical Report 1/2012, Instituto de Engenharia Biomédica / NNIG, December 2012. [3](#)
- [2] T. Amaral. Activities report from August 2012 to April 2013, April 2013. [3](#)
- [3] T. Amaral, J. Marques de Sá, L. M. Silva, L. A. Alexandre, and J. M. Santos. Improving performance on problems with few labelled data by reusing stacked auto-encoders. (Under review), 2013. [3](#)
- [4] T. Amaral, L. M. Silva, and L. A. Alexandre. Using different cost functions when pre-training stacked auto-encoders. Technical Report 1/2013, Instituto de Engenharia Biomédica / NNIG, April 2013. [3](#)
- [5] T. Amaral, L. M. Silva, L. A. Alexandre, C. Kandaswamy, J. M. Santos, and J. Marques de Sá. Using different cost functions to train stacked auto-encoders. In *Mexican International Conference on Artificial Intelligence (MICAI)*, pages 114–120, 2013. [3](#)
- [6] T. Amaral, L. M. Silva, L. A. Alexandre, J. Marques de Sá, and J. M. Santos. Transfer learning using rotated image data to improve deep neural network performance. (Under review), 2014. [3](#)