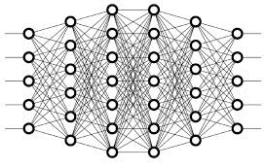
Statistical Learning with Deep Artificial Neural Networks

Task 2



Context and Data

Reading and diagnosing chest x-ray images may be a relatively simple task for radiologists but, in fact, it is a complex reasoning problem which often requires careful observation and knowledge of anatomical principles, physiology and pathology. Such factors increase the difficulty of developing a consistent and automated technique for reading chest X-ray images while simultaneously considering all common thoracic diseases.

The aim is to diagnosis chest X-ray images between normal and with effusion. The dataset we are going to use consists of 500 normal chest x-ray (folder named normal) and 500 with effusion x-ray (folder named effusion), taken and selected from the public NIH ChestXray14 dataset:

https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-available-chest-x-ray-datasets-scientific-community.

Data sets are available in two zip files, one each class: **normal.zip**, **effusion.zip**.

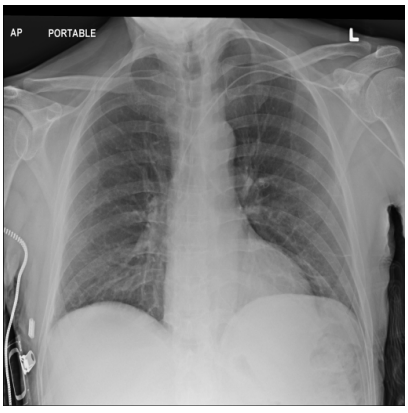


Figure 1: Normal chest x-ray of normal1.png file

Questions:

1. Separate the data into 2 sets: training set (2/3) and test set (1/3). Use this partition in the training phase (and validation phase if necessary) and the test phase of each of the sections that are presented below. Use the value `1234` as random seed to do the partition.
2. Implement a CNN architecture of your choice that contains at least 2 convolution layers.
3. Define conveniently the model (optimization, loss, metric, ...).
4. Using tfruns package tune the hyperparameter batch_size exploring the grid c(16, 32, 64)
5. Implement an early-stopping `callbacks()` to interrupt training when validation accuracy stops improving for more than two epochs.
6. Assess the performance of the CNN predicting the categories of test images and obtain the confusion matrix.
7. Implement a Convolutional Autoencoder (CAE), with 10 nodes in z layer (or bottleneck). Feel free to choose the number of convolutional layers, filter sizes, number of filters, ...
8. Represent graphically the results from images test to show the association between z layer activations and the class images.

Important remarks

- Answer the questions in a reasoned way, adding the necessary comments, not just only the code.
- A R markdown parameterized report as dynamic as you can.
- Use relative paths instead of absolute paths to read / write files, to make it easier to run the code outside of your computer.

Delivery / Deadline

A zip file including:

- the Rmd file used as template for the report,
- the output reports in pdf and/or html files.

Deadline: 18th of May, 2022