Deep Learning seminar Second deliverable:

Forecasting with Neural Networks

Sergi Bernet Andr´es

March 9, 2022

# Introduction

This document summarizes all the requirements for the second deliverable of the section: Deep Learning.

We’re working in a project with *Ajuntament de Barcelona*, and our objective is to predict weather in order to help the city to manage resources efficiently, and have a faster response in extreme situations such as heat/cold waves or even floods.

# Datasets

The *Data department* provided us two datasets to perform our job:

1. **Rainfall dataset**: Contains the monthly average from 1786 in Barcelona.
2. **Temperature dataset**: Same information type but the data is about **temperature** from 1780.

In addition, the environmental department also requires our help but for *air pollution*. This time, the data comes from ([https://opendata-ajuntament.barcelona.cat/data/en/dataset/qualitat-aire-detall-bcn)](https://opendata-ajuntament.barcelona.cat/data/en/dataset/qualitat-aire-detall-bcn) and requires some processing since the department is a bit lazy.

Nevertheless, the department is asking us to evaluate what happened during **Covid-19** to the air quality and prepare a model to ’predict’ if this will happen again.

# Objective

As first deliverable, there are transversal requirements:

1. **Analyze and prepare the data**: Analyze the data you’re gonna work with. Since this time we’re handling *temporal sequences* and you’ve seen a single class from Xavier Vilas´ıs, this analysis can be brief, but consider **trend**, **seasonality**, **stationarity** and all those concepts would be great.
2. **Create your model**: This time, two models are required: 1D-CNN and LSTM. Remember that preparing properly the data will help you a lot since you’ll need to process it a single time.
3. **Evaluate results**: Once you’re satisfied with your results (your criteria as data scientists) explain them and why are good (or not).
   1. **Using loss,**

Note that *rainfall* and *temperature* datasets are very straightforward to use. In both cases, predict the rain/temperature for Barcelona for this 2022.

Those of you who selected the *pollution* problem, here are some hints:

1

1. The dataset comes splitted in several .csv files. To obtain an initial approach, consider working with a single dataset (which means, single year data) or even smaller (trimesters) and predict the following trimester (for example, if you’re using the data from 2021, use the first 9 months for train and the last 3 to test).
2. Even this previous step reduces your samples, there’s data for all the measure stations of the city. I strongly recommend using a single contaminant and a single station such as:
   1. Contaminant: 8-di´oxido de nitr´ogeno / 7-Mon´oxido de nitr´ogeno / 9- 2.5 PPM / 10- 10 PPM.
   2. Station: 4-Poblenou / 42-Sants / 43-Eixample / Barri de Gr`acia) This way, you’ll downsample from 850k samples to 25k.

# Optative exercises

There are a couple of exercises defined below in order to obtain extra mark.

1. **Rainfall+Temperature**: Temperatures tend to decrease when rains since water drops fall from colder layers of the atmosphere (the gradient is about 5ºC for each km). In this optative exercise the objective is to analyze both time series and get some insights about this behavior. Is there a correlation? Explain what you’ve seen there.
2. **Anomaly detection**: Based on the last slide, perform anomaly detection to any of your models, explaining the steps and your results.

# Doubts, questions

Do not hesitate to contact me with your doubts at: sergi.bernet@salle.url.edu.

The deadline for this deliverable is the 28/03/2022. Note you have **2 weekends**, don’t wait until the end, there are more deliverables and handle all of them at the same is a lot of work to deal with.

2