REPORT ANNDL2021_Homework2

TEAM: Cooper

- Davide Canali
- Filippo Caliò
- Natalia Bagnoli

We started our work using a simple model with LSTM modules in series and a Dropout layer, followed by a Dense layer at the end of the model in order to predict the next value in the series and using only normalization technique for the preprocessing.

Noting that the model predicted the average and only a few other interesting parameters, we changed our model using an encoder-decoder architecture with 2 LSTMs as encoders and 2 LSTMs as decoders, which were initialized by the state of the respecting cell in the encoder. In this way all the layers were bidirectional.

As a result, keeping the last part unchanged (Dropout + Dense layers) and noting a significant improvement in the predictive capacity of the model, we added a Convolutional layer with the Max Pooling process in order to extract more information from the input data and improving the result even more.

Using this architecture, we added the standardization as pre-processing, and we tried different combinations of window, telescope and stride obtaining the following observations:

- With a too large window, the model only predicted the average like in the first model tried. On the other hand, with a too small window, the model did not have enough information to make accurate predictions.

- Performances improved using a telescope with a few tens of units compared to a telescope of 1 unit.

The final parameters chosen were the following:

- Window = 400
- Stride = 25
- Telescope = 20

Therefore, we tried to use more complex neural networks, using more LSTM or GRU layers, adding batch normalization after Convolution layer and before the Dense layer.
We also increased the convolutional part as well as the final dense layers in order to extract more information but, however, we do not obtain significant improvements.

We also tried to implement the Attention (like in seq2seq model) to try to improve the performance of the Encoder-Decoder: one part of the model sent queries to another part to create an Attention Vector to be passed back to the first part, but we were not very successful in doing this.

This presented below is the schema of our final model: