Machine Learning

Practical work 12 - Recurrent Neural Networks

Teachers: A. Perez-Uribe (Email: andres.perez-uribe@heig-vd.ch) & J. Hennebert
Assistants: Benoît Hohl (Email: benoit.hohl@heig-vd.ch) and Simon Walther (Email: simon.walther@heig-vd.ch)

Summary for the organization:

- Submit a report before Monday 11.12.2023 23h59 via Moodle.
- Modality: PDF report (max. 6 pages)
- The file name must contain the number of the practical work, followed by the names of the team members by alphabetical order, for example 13_dupont_muller_smith.pdf.
- Put also the name of the team members in the body of the report.
- Only one submission per team.

0. Notebooks and libraries

Download the notebook material from the Moodle platform and the databases.

1. Types of sequence-prediction problems

Jason Brownlee on the "Machine learning mastery" website presents a pretty simple examples of the 5 types of sequence prediction tasks that can be solved with Recurrent Neural Networks. Carefully read this page and follow the examples in Python.

https://machinelearningmastery.com/sequence-prediction-problems-learning-lstm-recurrent-neural-networks/

Nothing to include in the report regarding this point.

2. Time-series forecasting tutorial online

Follow the tutorial on "Time-series forecasting" by Jason Brownlee on the "Machine learning mastery" website. You will learn how to prepare data, develop, and evaluate an

LSTM recurrent neural network for time series forecasting. It is a fairly simple example that uses a small database to start with LSTM recurrent neural networks.

https://machinelearningmastery.com/time-series-forecasting-long-short-term-memory-network-python/

Nothing to include in the report regarding this point.

3. Race time prediction

The purpose of this exercise is to predict the following values of a sequence of multidimensional data. The objective is to train an LSTM network in the task of predicting the speed of a runner for a given slope by using her/his previous speeds during a given race as well as the slope profile of that part of the race. Finally, we can compute the expected time for a runner and compare it with the real performance.

You are provided with a database of several running trials including a certain number of measurements during each race. Each observation consists of time, elevation, distance from start, speed and slope.

Work to do and report

1. Explore different number of LSTM units, different lengths of previous data (sequence length) and training epochs. Show the configuration that performed the best. Observe the resulting complexity of the network (e.g., number of trainable parameters)

2. What is the largest error (speed prediction) you observed? Do you observe that most of those large errors show up for high speeds ? or low speeds? Why?

3. Using the predicted speeds for a given race, compute the expected time for a race and compute the difference between the real race time and the predicted race time in minutes. Provide the code of the cell that computes this prediction error.