MSE | MASTER OF SCIENCE IN ENGINEERING

HES-SO

Machine Learning

Practical work 12 - Autoencoders

Teachers: A. Perez-Uribe (Email: andres.perez-uribe@heig-vd.ch) & J. Hennebert
Assistants: Benoît Hohl (Email: benoit.hohl@heig-vd.ch) and Yasaman Izadmehr (Email: yasaman.izadmehr@heig-vd.ch)

Summary for the organization:

- Submit a report before Tuesday 13.12.22 11h00 via Moodle.
- Modality: PDF report
- The file name must contain the number of the practical work, followed by the names of the team members by alphabetical order, for example 12_dupont_muller_smith.pdf.
- Put also the name of the team members in the body of the report.
- Only one submission per team.

1. Autoencoders: denoising application

This exercise has the objective of using a convolutional autoencoder in the task of denoising. The idea is to train an autoencoder with digits of the MNIST database. The encoder part is supposed to find a reduced representation of the images while the generator part should be capable of generating the input image from the reduced representation (latent space). When a noisy input is presented to the autoencoder, the encoder captures the essence of the object in the image (i.e., its features) and reconstructs an image without noise. Here, we will explore if the auto encoder can remove different amounts of synthetic noise applied to the original images.

2. Autoencoders: data augmentation application

This exercise has the objective of using a convolutional autoencoder to learn how to generate a particular object, for example, a digit. We will use the images corresponding to a particular digit, let's ay, the "2", from the MNIST database to train our autoencoder. We will test if using a limited amount of examples (e.g., 1000, or 100 or even only 10) can be enough to generate multiple new instances of that object.
Finally, we will test if a pre-trained classifier is able to recognize those new instances of the object we have generated with the autoencoder.

3. Autoencoders for anomaly detection

This exercise has the objective of using an autoencoder to learn a “normal” data distribution to subsequently spot outliers. In this particular case study, we will use a set of time-series that correspond to a normal situation (e.g., data from a healthy person, or a normally functioning machine) to train an auto-encoder. We will consequently use our pre-trained model as an anomaly detector. To do this, we try to predict the values of a very similar data set but containing outliers.

Report

DENOISING

- 1. Provide a screenshot of the code you wrote to create and train the convolutional autoencoder for denoising digits from the MNIST database.
- 2. Provide results showing some examples of the ten digits of the MNIST database before and after denoising for the three different levels of noise: noise factor = 0.5, 0.7, and 0.9
- 3. Conclude based on the obtained results

DATA AUGMENTATION

- 4. Generate different synthetic digits (e.g., twos, threes and eights). Comment about the diversity of the synthetic objects the autoencoder is generating. Is it possible to measure that diversity ?
- 5. Provide a summary of your experiments using the autoencoder for data augmentation and comment your results (e.g., modify the latent variables of an autoencoder that was trained with only 10 examples such that it generates digits that are no longer recognized by the pre-trained classifier).

ANOMALY DETECTION

- 6. Answer questions 1 and 2 in the notebook.
- 7. Perform exercise 1. Comment on your process : Compare the different models you tried. What did you do to improve the models ? What is improved ? Why do you think it works ? Is it working better with a convolutional autoencoder ?