Variational Autoencoders

Autoencoders are neural networks. They are unsupervised machine learning algorithms, which estimates the output values to be equal to the inputs from reduced data representations. Therefore, autoencoders are used to reduce the dimensionality of input data into smaller representation spaces. Moreover, original data can be reconstructed from the compressed reduced representations.

There exist other popular machine learning algorithms that are used to extract reduced data representation, such as Principal Component Analysis (PCA), which calculates a linear projection of the original data into a reduced subspace. However, autoencoders can go beyond, being able to learn non-linear transformations between the input data and the compressed representation space, which can be applied in much more general problems. For instance, it can be more efficient to use an autoencoder with several layers for a large input data space rather than a huge linear transformation with PCA.

There are different types of autoencoders. They can be just made of a simple hidden layer with an input and an output layers. The hidden layer represents the compressed or reduced latent representations, also called the bottleneck. More complex autoencoders can be made of several deep layers, or even convolutional neural network (CNN) layers, which can be more adequate for image, video or data series. For instance, autoencoders based on deep fully connected layers does not take into account the fact that a signal can be modelled as a sum of a set of signals. However, CNN autoencoders use the convolution operator to exploit that fact.

Autoencoders can be used as pre-trained layers from other neural networks (NN) models to apply transfer learning to enhance the encoder/decoder of other problems. Typical applications of autoencoders are data compression, grey level image colorization, image denoising, image super-resolution, watermark image removal, …

An autoencoder network consists of three different parts:

- Encoder: This part is a set of layers that compresses the input into a reduced latent space representation.
- Code: It represents the compressed representation of the input, which is further used as the input to the decoder.
- Decoder: This part is symmetrical with respect to the encoder. It decodes the encoded data representation to the original data representation space. The decoded data is a lossy reconstruction of the original data, which is reconstructed from the latent representation.

Variational AutoEncoders (VAE) are a slightly different and interesting way of autoencoding. In the last few years, deep learning based generative models have gained more interest. VAEs are a kind of NN generative model. The term generative model stands for statistical models able to generate new instances from a learned joint data and target probability distribution. That is, you can generate new data samples from this latent probability distribution.

Recent deep generative models have shown very successful to produce highly realistic images or videos. Two out of the most important families of deep generative models are Generative Adversarial Networks (GANs) and VAEs.

VAEs are autoencoders whose latent spaces are encoding probability distributions, represented by a set of random variables, allowing the decoder to generate new data instances. That is, VAEs are not deterministic functions. Moreover, the term “variational”

stands for the variational inference method used in statistics, which is the one applied to learn the parameters of the latent space in VAEs.

In a few words, VAEs work in the following way. An encoder network turns the input samples into parameters of a probability distribution in a latent space. Then, given an input data sample, a sample is randomly generated from the distribution represented by the latent space. Finally, a decoder network maps these latent space samples back to the original input data space, producing a new data sample.

The parameters of a VAE model are typically trained with a loss function based on a reconstruction loss forcing the decoded samples to match as much as possible the initial inputs, and a probabilistic divergence measure between the learned latent distribution and the prior distribution, which plays the role of a regularization term.

This lecture will address fundamental questions to understand autoencoders and their main properties, in particular, about VAEs, such as their statistical foundation and the way these generative models have been implemented in a NN architecture. The lecture will be also include some practical examples and exercises to understand how to put into practice VAEs and the other basic types of deterministic autoencoders.

The outline of the lecture is as follows:

- Auto-encoders:
    - definition and structure,
    - linear vs non-linear embedding.
    - deep and convolutional autoencoders
- Generative NNs. Modelling uncertainty:
    - hidden random variable model of observed data.
- Probabilistic inference:
    - variational inference.
- From deterministic to variational auto-encoders:
    - latent space distribution
    - the re-parametrization trick.
- Practical applications in imaging.
- Practical examples and exercises (Keras + Google-Colab)