Practical Work 13 – 23/05/2024
Generative Models

Objectives

The objective of this PW is to experiment with generative AEs VAEs and GANs.

Submission

— **Deadline** : Wednesday, 5 June, 12am (noon)

— **Format** : Zip file with your jupyter notebooks.

Exercise 1 Autoencoders and Variational Autoencoders

Plain Autoencoders

Implement different versions of a "plain" auto-encoder with 128 latent dimensions.

a) Model with only fully connected layers,

b) Model with only CNNs.

Train them with Fashion-MNIST. Track the reconstruction error (train and test per epoch).

What model leads to the best (smallest) reconstruction error on the test set ? Verify that also by looking at individual samples from the (test) set.

Explore the latent space of the trained models by constructing images from latent variables as follows :

a) Pick a random image and, by encoding it, map it to the latent space (variable z). Then look at the resulting reconstructions from scaled z-variables, i.e. $\tilde{z}(t) = t\, z, t > 0$.

b) Linearly interpolate between latent variables associated with encodings of two arbitrarily selected images and reconstructing them.

c) Reconstruct from randomly sampled latent variables, sampled from a d-dim standard normal distribution.

Hint : For the CNNs you may first start with the exercise with GANs below.

Variational Autoencoder

Transform your CNN-only AE from the previous subsection into a VAE, again using 128 latent dimensions. During training track both, the reconstruction loss and the KL-loss. Could you improve the reconstruction error by comparing with the plain autoencoder ?

Explore the latent space of the trained VAE - following the same steps as above. Describe and interpret what you observe. How does the random generation by sampling from the latent space work now ?

Exercise 2 Generative Adversarial Networks : DCGAN (optional)

Start from the DCGAN tutorial (for tensorflow or pytorch).

a) Go in detail through the models for the Generator and Discriminator, compute the dimensions of the activation maps for each of the layers - given the filter size, padding and in/out channels. Check how the tensor shapes are transformed into the shape in latent space.
b) Then, adjust the code to make it work with Fashion-MNIST and train the GAN.
c) Keep track of the generator and the discriminator loss. Can you observe stable behaviour ?
d) Periodically, create a mosaic plot (with 8x8 images) generated from the same randomly sampled latent variables. This allows you to visually track progress of how the training makes progress in learning the distribution and how to generate new FashionMNIST-like images.
e) Play with the hyper parameters (e.g. learning rate) to see whether you can see yet the same stable behaviour.

Exercise 3 Optional : Review Questions

a) Explain the difference between an auto encoder and a variational autoencoder.
b) What are the two components of the loss of a variational autoencoder and what do they mean.
c) Given the same encoder and decoder network structure : Does an AE or a VAE (sufficiently long trained) provide the smaller reconstruction error ?
d) Explain the basic concept of how GANs work.
e) Why are GANs considered less robust to be trained ?
f) Explain the basic concept of how diffusion models work. What is learned ?
g) What is the underlying principle that defines the loss when training diffusion models
h) Explain why diffusion models are more robust to be trained than GANs and why are diffusion models computationally expensive when generating new examples.