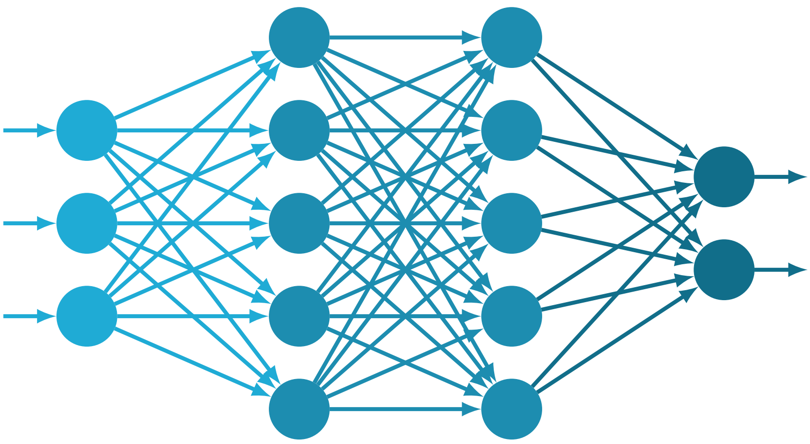
KU LEUVEN

Artificial Neural Networks & Deep Learning

Exercise Session 4

Prof. Dr. ir. Johan A. K. Suykens

Assistants
Sonny Achten
Bram De Cooman
David Winant
Xinjie Zeng



2023-2024

SESSION 4

Generative Models

Note: There is no need to repeat the information provided during the guided session. Answer the questions as concise and accurate as possible. Only plot what is relevant to your claims, not necessarily everything that you produce. The *page limit* for this exercise is *4 printed pages*, including figures.

This assignment will revolve around the exploration and use of various generative models. Colab links to all notebooks are provided, and they can also be found on Toledo.

4.1 ENERGY-BASED MODELS

This section will investigate the use of probabilistic models based on the use of an energy function. In particular, we will focus on *Restricted Boltzmann Machines* and its deep extension. Please use the `rbm.ipynb` and `dbm.ipynb` notebooks for these exercises.

Restricted Boltzmann Machines

We consider the notebook `rbm.ipynb`. Please answer the following questions.

1. In the notebook, the training algorithm refers to the *pseudo-likelihood*. Why is that? What are the consequences regarding the training procedure?
2. What is the role of the number of components, learning rate and number of iterations on the performance? You can also evaluate it visually by reconstructing unseen test images.
3. Change the number of Gibbs sampling steps. Can you explain the result?
4. Use the RBM to reconstruct missing parts of images[1]. Discuss the results.
5. What is the effect of removing more rows in the image on the ability of the network to reconstruct? What if you remove rows on different locations (top, middle...)?

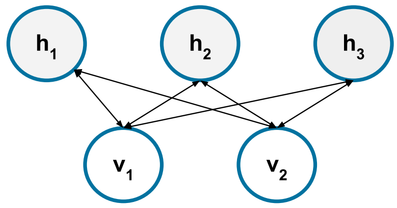


Figure 4.1: Schematic representation of a Restricted Boltzmann Machine.

[1] An interesting video about image reconstruction using RBM's can be found at https://www.youtube.com/watch?v=tk9FTdKOL5Q.

Deep Boltzmann Machines

Now consider the notebook `dbm.ipynb` and go trough the code. Afterwards, answer the following questions.

1. Load the pre-trained DBM that is trained on the MNIST database. Show the filters (interconnection weights) extracted from the previously trained RBM (*cf. supra*) and the DBM, what is the difference? Can you explain the difference between filters of the first and second layer of the DBM?

2. Sample new images from the DBM. Is the quality better than the RBM from the previous exercise? Explain.

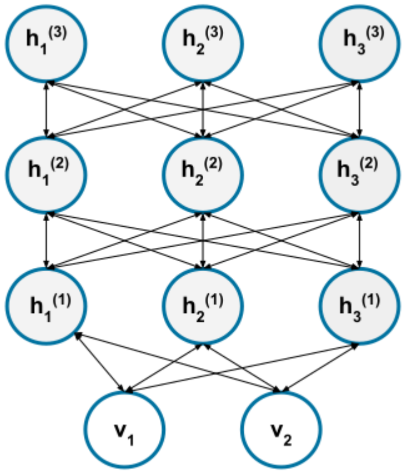


Figure 4.2: Schematic representation of a 3 layer Deep Boltzmann Machine.

4.2 GENERATOR AND DISCRIMINATOR IN THE RING

This exercise revolves around the use of *Generative Adversarial Networks*, where a discriminator and a generator compete against each other. Use the corresponding `gan.ipynb` notebook and make sure that you train the model long enough, such that it is able to generate convincing images. Monitor the loss and accuracy of the generator and the discriminator.

1. Explain the different losses and results in the context of the GAN framework.

2. What would you expect if the discriminator performs proportionally much better than the generator?

3. Discuss and illustrate the convergence and stability of GANs.

4. Explore the latent space and discuss.

5. Try the CNN-based backbone[2] and discuss.

6. What are the advantages and disadvantages of GAN-models compared to other generative models, *e.g.* the auto-encoder family or diffusion models? Think about the conceptual aspects, the quality of the results, the training considerations, *etc*.

[2] A model is often comprised of a general architecture, *i.e.* a GAN, a VAE, *etc.* The specific networks inside, with their specific type and number of layers, neurons, *etc.*, is called a *backbone*.

4.3 AN AUTO-ENCODER WITH A TOUCH

This section will revolve around the training and use of *Variational Auto-Encoders* (VAEs). Use the corresponding notebook `vae.ipynb` can be found on Toledo. To ease the training, we consider a fully connected network as backbone.

1. In practice, the model does not maximize the log-likelihood but another metric. Which one? Why is that and how does it work?
2. In particular, what similarities and differences do you see when compared with stacked auto-encoder from the previous assignment? What is the metric for the reconstruction error in each case?
3. Explore the latent space using the provided code and discuss what you observe.
4. Compare the generation mechanism to GANs. You may optionally want to consider similar backbones for a fair comparison. What are the advantages and disadvantages?

4.4 OPTIONAL: GENERATION AND CREATIVITY

In this section, the goal is to create a practical generative model from scratch. Don't hesitate to base yourself on the codes given in the previous sections. Choose a dataset of your liking, a model and train it. Discuss all your design and training choices thoroughly as well as the obtained results.

Note: Maybe more than for other AI applications, generative models can be subject to particularly deeper ethical issues. Pay attention to the license of your dataset and to the ethical considerations of the obtained model. Please avoid faces, other highly person-related cases, possible military or surveillance applications...