

# Artificial Neural Networks: Exercise session 4

Stijn Staring (r0620003)

May 22, 2021

## 1 Restricted Boltzmann Machines

The Boltzmann machine is a parametrized generative model representing a probability distribution. It is a neural network that belongs to Energy based models. A Boltzmann Machine consists out of a layer of visible units  $\mathbf{v}$  and hidden units  $\mathbf{h}$ . The standard type of BM uses binary-valued hidden and visible units. The difference between a Restricted Boltzmann Machine and a conventional Boltzmann Machine, is that the units of one layer are not connected with each other. Because of this restriction, the partition function in the joint probability seen in Eq. ?? becomes tractable. Training an RBM means adjusting the RBM weight values and biases such that the probability distribution Eq. ?? fits the training data as well as possible.

An RBM trains in an unsupervised manner.

$$E(\mathbf{v}, \mathbf{h}) = -\mathbf{v}^T \mathbf{W} \mathbf{h} - \mathbf{b}^T \mathbf{v} - \mathbf{a}^T \mathbf{h}, \quad (1a)$$

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})}. \quad (1b)$$

Eq. 1

Training is done by making updates by calculating gradient with the log likelihood.

Learning rate/Iterations	5	10	20	Iterations/Components	5	10	20
0,005	141	134	127,41	5	185	155,08	134,46
0,01	166	147,66	125,55	10	186	154	130
0,05	141,15	134,35	127,41				
0,1	134,46	130,24	130,81				

Table 1: Absolute value of pseudo-likelihood on the training data after training.

- The small hyperparameter tuning in Table 1 shows that the smallest
  - when performing the Gibbs sampling the more samples are done the further from the original test image (logically, because after a certain amount of samples noise will only be increased).
  - In ideal world, would not need Gibbs sampling but because calculating the expectation of the model distribution is intractable, have to need Gibbs sampling to calculate a sample of the model distribution. From there the heuristic can be used for the maximum log likelihood and to calculate updates for the weights and biases.
- It is expected that when more iterations of the Gibbs sampling are performed, a sample is retrieved that better approximates one coming from the model distribution.
- Property of Gibbs sampling is that only after an infinite amount of iterations it can be guaranteed that the generated sample originates from the probability distribution tried to be sampled from.