Restricted Boltzmann Machine - Homework 2

Thibault Desjonquères 19990602-8890

October 7, 2022

Code explanation

Code description.
First, the training of the RBM takes place, using the 4 patterns of the XOR function. Second, the RBM is tested by using the 8 configurations of the boolean function with 3 variables. In order to asses the quality of the RBM, the Kullback-Leibler divergence DK_L is calculated as follow :

$$DK_L = \sum_{\mu=1}^{p} P_{data}(\mathbf{x}^{(\mu)}) log(\frac{P_{data}(\mathbf{x}^{(\mu)})}{P_B(\mathbf{s} = \mathbf{x}^{(\mu)})}) \tag{1}$$

The probability of using one of the XOR pattern is set at $P_{data}(\mathbf{x}^{(\mu)}) = 1/4$. The term DK_L quantifies how well does the RBM manage to reproduce a data distribution. The lower the DK_L, the better the performance of the RBM. The parameters used to obtain this plot are given in figure 1.

Results discussion

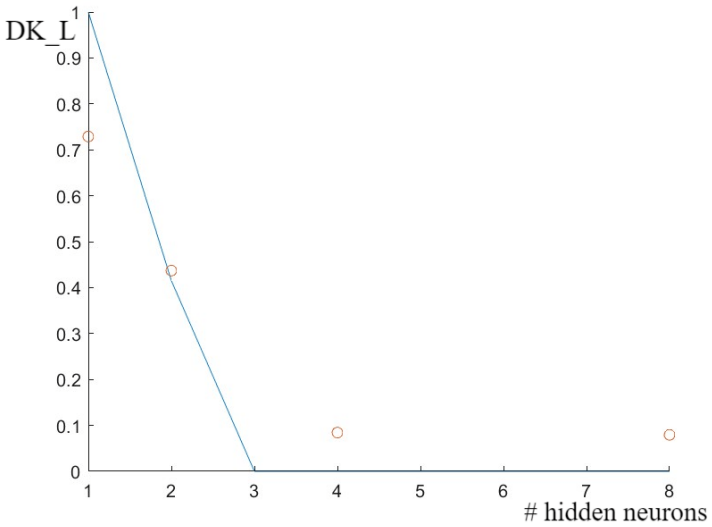

Figure 1: Theoretical DK_L curve and experimental DK_L as a function of the number of hidden neurons.

Training	
nTrials	1000
nInner	200
nOutter	100
eta	0.008
Validation	
nInner	300
nOutter	300
Results	
M=1	0.7289
M=2	0.4370
M=4	0.0844
M=8	0.0795

Table 1: Parameters used to obtain the experimental results shown in figure(1)

.

	[-1,-1,-1]	**[1,-1,1]**	**[-1,1,1]**	**[1,1,-1]**
M=1	0.1767	0.0667	0.0700	0.2567
M=2	0.0733	0.1567	0.2433	0.2433
M=4	0.2267	0.2833	0.2067	0.2100
M=8	0.2433	0.1667	0.2533	0.2767

Table 2: Probability of each of the indicated pattern to be given as outputs, for the respective number of hidden neurons.

.

With more hidden neurons, DK_L converges to zero. This is represented by both the theoretical curve and the experimental one. The code took about 2 hours to run. Obtaining this curve required running my code for one hour. The longer I trained my network, the better the results, therefore I suppose this result could be improved by training the network even more. I do not understand the result obtained for 1 hidden neuron. It appears that my experimental result performs better than what the theory predicts, which is odd. However, the convergence to 0 is irrefutable. The more neurons, the more decision boundaries can be drawn and the better the accuracy. The probability of appearance for each of the indicated pattern is shown in figure 2. With more hidden neurons, this probability fluctuates around 25%. The rest of the time, an other pattern of the Boolean function in 3 variables comes as an output.