# Restricted Boltzmann Machine

The *learning rate* was set to 0.004, $k$ to 10, *batch size* to 100, and training was performed over 1500 *epochs*. These parameters offered a good balance across all hidden neuron configurations. The number of *epochs* ensured M=4,8 accurately reproduced the target distribution, while $k$ was selected to guarantee the system "forgot" prior states and avoided dependence on earlier dynamics. Reducing $k$ degraded the quality of distribution approximation. The *batch size* allowed for an appropriate representation of the target training distribution, approximately [1/4, 0, 0, 1/4, 0, 1/4, 1/4, 0].

For sampling, I initialized the visible state randomly, ran the dynamics for $k$ steps, sampled the updated states by incrementing the position corresponding to the new state, and repeated this process 10,000 times. Afterwards, I normalized the array of the counted frequencies, and computed the Kullback-Leibler divergence, as illustrated in Figure 1. The results validate the accuracy of the reproduced distribution, as shown in Table 1. Furthermore, when studying Table 1 and Figure 1, one can conclude that the XOR function is difficult to learn, and that the number of hidden neurons has to be at least more than two to get a precise result. One can also see that eight hidden neurons are not necessarily better than four, in this case.

| M (Hidden Neurons) | $D_{KL}$ | $D_{KL}$-bound |
|---|---|---|
| 1 | 0.58433 | 0.69315 |
| 2 | 0.36590 | 0.34657 |
| 4 | 0.00312 | 0.0 |
| 8 | 0.00339 | 0.0 |

Table 1: Comparison of $D_{KL}$ and $D_{KL}$-bound for different values of M (hidden neurons). The plot of these values is represented in Figure 1.

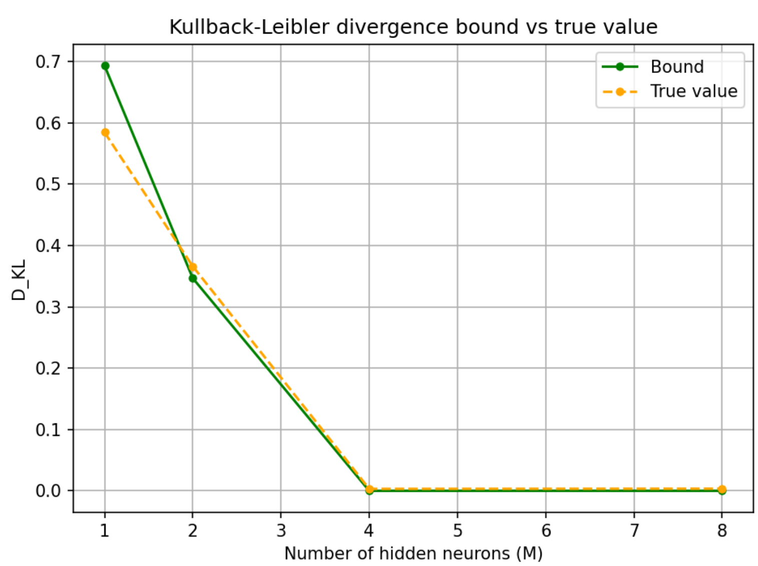


Figure 1: The figure represents the Kullback-Leibler divergence ($D_{KL}$) as a function of hidden neurons (M). The estimated upper bound (Green), and the value calculated after sampling (Orange).