Homework 2 Restricted Boltzmann machine

In my training program, the CD-k will iterate for 1000 times with $k = 500$ and mini-batches with 20 randomly selected patters. In practice it's long enough for the errors to converge.

In calculating Kullback-Leibler divergence, 3000 patterns are fed into the trained neural network then update the visible and hidden neurons back and forth for 2000 rounds. Accumulate the times that each pattern shows up, then this will be used to calculate the distribution that Boltzmann machine approximates, P_B.
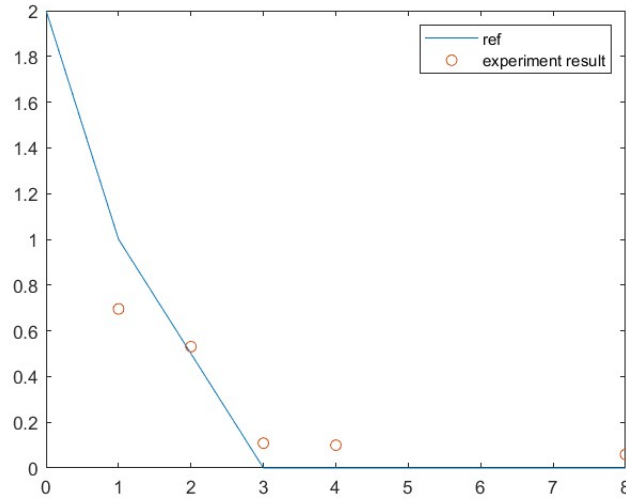

Figure 1: Numerical results

The figure above illustrates how the Kullback-Leibler divergence changes according to different number of hidden neurons in a restricted Boltzmann machine. The experiment result D_{KL} of $M = 1, 2$ are smaller than the corresponding upper bound which means it follows the equation below. However, for $M = 3, 4, 8$, the result is slightly larger than the theory. The possible reason could be that the CD-k algorithm is not an optimal method. And the learning rate η is still quite large for 20 patterns in a minibatch.

$$D_{\mathrm{KL}} \leq \log 2 \begin{cases} N - \lfloor \log_2(M+1) \rfloor - \frac{M+1}{2^{\lfloor \log_2(M+1) \rfloor}} & M < 2^{N-1} - 1, \\ 0 & M \geq 2^{N-1} - 1 \end{cases}$$

Small k or large η will result in the model not converge. The corresponding D_{KL} will so that not stay below the boundary.