Homework 2 Restricted Boltzmann machine

In my training program, the CD-k will iterate 1000 times with $k = 500$ and mini-batches with 20 randomly selected patterns. $\eta = 0.003$. The training process will repeat for 5 times to find the minimal D_{KL}.

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 the Boltzmann machine approximates, P_B.

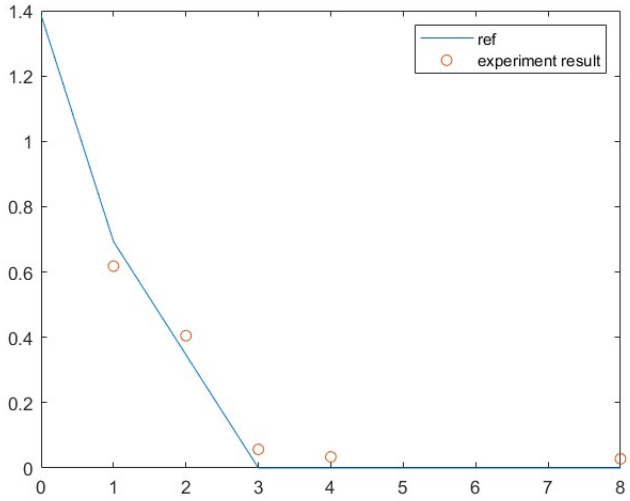

Figure 1: Numerical results

The figure above illustrates how the Kullback-Leibler divergence changes according to different numbers of hidden neurons in a restricted Boltzmann machine. The experiment result is roughly above the upper boundary. It could be because the CD-k algorithm is not an optimal method. And since calculating D_{KL} is a stochastic process, the value can be more precise under a large number of tests.

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

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