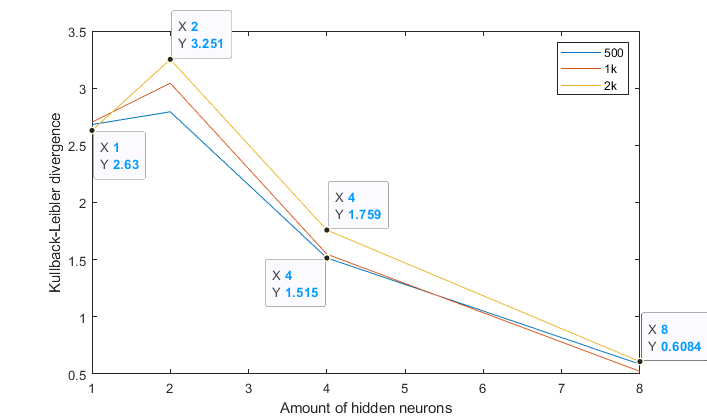
A restricted Boltzmann machine (RBM) was trained using 3 visible and *M* = 1,2,4,8 hidden neurons using the CD-k algorithm using *k* = 100, a learning rate of η = 0.1 and 100 training steps. To compute the Kullback-Leibler divergence I iterated the dynamics of the RBM after training it using the parameters mentioned and counted the frequency at which the different patterns occur. This was then done for all the different amount of hidden neurons *M* and the Kullback-Leibler (KL) divergence was then plotted as a function of the number of hidden neurons as can be seen in Figure 1. The KL divergence plotted in Figure 1 is an average of 50 independent trainings to understand what divergence is to be expected at that number of hidden neurons. 50 times is quite low and could allow for some error and a better amount would be about 500 times as done in the lectures notes.

Figure 1. The KL divergence for different number of hidden neurons on an average of 50 trainings.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Nh | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 |
| 1 | 0.1221 | 0.1386 | 0.1461 | 0.1444 | 0.0987 | 0.1143 | 0.1120 | 0.1238 |
| 2 | 0.1418 | 0.1629 | 0.1553 | 0.1598 | 0.1101 | 0.0863 | 0.0967 | 0.0872 |
| 4 | 0.2519 | 0.2256 | 0.2455 | 0.2219 | 0.0147 | 0.0132 | 0.0133 | 0.0140 |
| 8 | 0.2548 | 0.2480 | 0.2366 | 0.2476 | 0.0028 | 0.0046 | 0.0022 | 0.0035 |

Table 1. Percentage of patterns converged to the different patterns (x1-4 is the trained patterns) using 2000 samples (yellow line), for different number of hidden neurons.

As can be seen in Table 1