Report of FFR135 HM2.2: RBM with XOR dataset

Siyu Hu, gushusii@student.gu.se, ID: 19950910-3702

2024-10-03

Running the program 20 times, the values of KL for number of hidden neurons M = 1, 2, 4, 8 shows in Figure 1. The red dashed line represents the theoretical KL divergence (Equation 4.40).

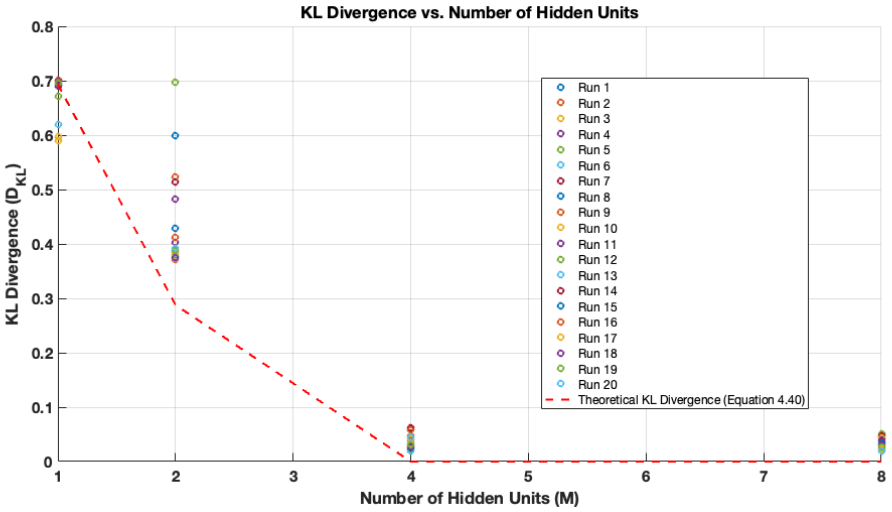

Figure 1: KL divergence value with 20 runs, other parameters: learning rate = 0.001; CD-k steps k = 10; epochs = 1000; batch size = 100; number of generated samples = 100000

Since KL divergence value stands for the difference between RBM model generated samples probability with the target probability, the KL value suggests whether the RBM learns all the information (including target probability) in the XOR dataset.
For $M = 1$: The KL divergence is high (around 0.6-0.7). **It means that the RBM struggles to learn the XOR dataset with only 1 hidden unit.**
For $M = 2$: The KL divergence decreases significantly (to around 0.4-0.5). **It shows that with one more neurons, RBM can learn the XOR distribution better.**
For $M = 4$: The KL divergence approaches zero. **It means that RBM with 4 hidden neurons can accurately learn the XOR dataset distribution and shows near-perfect performance.**
For $M = 8$: The KL divergence remains near zero, with no additional improvement compared to $M = 4$, **indicating that using four more neurons can not significantly improve the performance of RBM, and 4 hidden neurons are sufficient for the task**