Introduction

Boltzmann Machines

A Boltzmann Machine consists of a set of binary units, s and all these units are connected to each other with a weight, w_{ij} associated with each connection. The global energy, E of the Boltzmann Machine is given by:

$$E = - \left(\sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i \right)$$

While learning from data, we try to adjust the weight parameters such that the stationary distribution of the Boltzmann machine closly represents our dataset. For determining the stationary distribution of the Boltzmann Machine we start with random states for all the binary units and iterate over them setting them to $+1$ with a probability of $\frac{1}{1+e^{-2a_i}}$, -1 otherwise. After a few iterations the machine converges to its stationary distribution.

Now for comparing this stationary distribution to our dataset we generate some samples from the Boltzmann Machine and using that we compute the free expections using:

$$< s_i >= \sum_s s_i p(s), \quad < s_i s_j >= \sum_s s_i s_j p(s)$$

Similarly we compute clamped expectations from our dataset using:

$$< s_i >_c = \frac{1}{P} \sum_\mu s_i^\mu, \quad < s_i s_j >_c = \frac{1}{P} \sum_\mu s_i^\mu s_j^\mu$$

and then we update our weights based on these values:

$$\begin{aligned}
w_{ij}(t+1) &= w_{ij}(t) + \eta \frac{\partial L}{\partial w_{ij}} \\
\theta_i(t+1) &= \theta_i(t) + \eta \frac{\partial L}{\partial \theta_i} \\
\frac{\partial L}{\partial \theta_i} &= < s_i >_c - < s_i > \\
\frac{\partial L}{\partial w_{ij}} &= < s_i s_j >_c - < s_i s_j >
\end{aligned} \tag{1}$$

Mean Field Theory

Research Questions

Adding noise to the MNIST Dataset

TODO: Why do we need noise. We studied the accuracy of the Boltzmann Machine on different levels of noise.

Sampling vs Mean Field

Since the total possible number of states in a Boltzmann Machine is 2^n, therefore it's not feasible to use exact method for computing the free statistics. Therefore, we used sampling methods and Mean Field and compared the results.

Results

Accuracy of the model with varying levels of noise

0.1 Comparision of Sampling Methods and Mean Field Approximation

Conclusion

Appendix