**Boltzmann machine learning exercise**

In this exercise you implement the Boltzmann machine learning in the absence of hidden units. Write a computer program to implement the Boltzmann machine (BM) learning rule (see slides). The gradients are expressed in terms of the statistics of the data $\langle s_i \rangle_c, \left\langle s_i s_j \right\rangle_c$ and the statistics $\langle s_i \rangle, \left\langle s_i s_j \right\rangle$ of the model. The former can be easily computed from the data. The latter depend on the model and change during the learning. They need to be recomputed in each learning iteration. The computation of these statistics is intractable for large models.

In this exercise you are asked to implement the BM learning algorithm for small problems using exact computation and for large problems using MC sampling and mean field approximations.

- For small models (up to 20 spins) the computation can be done exactly. Make a toy problem by generating a random data set with 10-20 spins. Define as convergence criterion that the change of the parameters of the BM is less than $1 \times 10^{-13}$. Demonstrate the convergence of the BM learning rule. Show plot of the convergence of the likelihood over learning iterations.

- Apply the exact algorithm to 10 randomly selected neurons from the 160 neurons of the salamaner retina, as discussed in [Schneidman et al., 2006]. The original data file has dimension $160 \times 283041$, which are 297 repeated experiments, each of which has 953 time points. Use only one of these repeats for training the BM, ie. your data file for training has dimension $10 \times 953$. Reproduce [Schneidman et al., 2006] fig 2a.

- For larger problems, implement a Metropolis Hasting sampling method using single spin flips to estimate the free statistics $\langle s_i \rangle, \left\langle s_i s_j \right\rangle$ in each learning step. Produce a plot of the likelihood over learning iterations that compares the accuracy of your sampled gradient with the exact evaluation of the gradient for small systems. Investigate how many Monte Carlo samples are required so that the gradients are sufficiently accurate for the BM learning. Since the gradient is not exact and fluctuates from iteration to iteration, a convergence criterion is less straightforward. Propose a convergence criterion.

- Repeat the previous step, where you replace the MH sampling method by the mean field + linear response method to estimate the free statistics $\langle s_i \rangle, \left\langle s_i s_j \right\rangle$. For a small toy data problem ($n = 10 - 20$), produce a plot where you compare the likelihood versus learning iteration for the MH and MF approximations with the exact learning method.

- When you are convinced of the accuracy of your MH and MF learning algorithms for the BM, apply them to learn the connectivity between all 160 neurons of the salamaner retina data set. Produce plot of likelihood versus iteration for MF and MH learning.

- A much faster alternative method is to solve for the $w_{ij}, \theta_i$ directly from the fixed point equations $\langle s_i \rangle = \langle s_i \rangle_c$ and $\left\langle s_i s_j \right\rangle = \left\langle s_i s_j \right\rangle_c$ in the mean field and linear respons approximation (see slides). Possible complication may arise when the matrix $C$ is not invertible. Or, $C$ may be invertible, but the solution for $w$ has very large weights so that the MF approximation is very poor. Therefore, add a small diagonal $C \to C + \epsilon 1$ so that the solution has not too large weights. Plot the likelihood of the solution as a function of $\epsilon$ for the salamander data.

# References

[Schneidman et al., 2006] Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012.