Exercise 11: Hopfield Networks

A Hopfield network is a form of recurrent artificial neural network invented by John Hopfield in 1982. One example with seven units is represented in the figure. Let $\boldsymbol{W}$ denote the weight matrix (the matrix in which the connection strengths between the neurons are stored) of the Hopfield network and w_{ij} its elements.

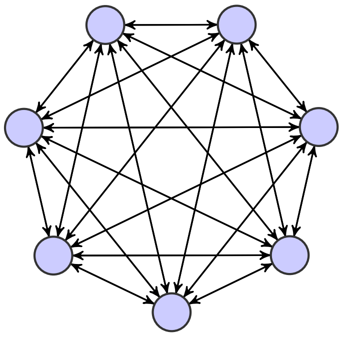

Figure 1: A Hopfield network of 7 neurons. Vertices represent neurons, edges indicate connections.

1. What are the symmetries and zero-entries of the weight matrix in a Hopfield Network?

In a Hopfield network, at each time step, each neuron computes an output based on its inputs. Since there is feedback, the activation does not stop after one step. But the state of the network (the value of all the units) changes in time. We are particularly interested in finding stable states, where the network does not change anymore.

We now consider a Hopfield network of four neurons and the set of outputs {–1,1}. We want to learn 2 patterns: (–1,1,1,–1) and (–1,1,–1,1). That means we want that when neuron one outputs –1, neuron two outputs 1, neuron three outputs 1 and neuron four outputs –1, and they keep these values at each step. Similarly for the second pattern. Assume zero threshold. The values selected for inactive/active states are the ones to be used in the calculation of the weighted sum.

2. What conditions does the existence of these patterns impose on the weight matrix? (What should a given neuron do for inputs that correspond to these patterns?) Give a concrete example of weights w_{ij} that fulfill these conditions.

Hopfield networks serve as content-addressable memory systems with binary threshold units. In the retrieval phase an N-dimensional vector $\mathbf{x}$, let's call it the input, is imposed on the network as its state. Typically, it represents a noisy version of a fundamental memory. Information retrieval is then done in accordance with the dynamical update rule. Consider a three unit Hopfield network with weight matrix:

$$W = \frac{1}{3} \cdot \begin{pmatrix} 0 & -2 & 2 \\ -2 & 0 & -2 \\ 2 & -2 & 0 \end{pmatrix}$$

3. Having three neurons how many different possible states can this network have?
4. Calculate for all possible states, what are the possible following states. Assume asynchronous updates (*i.e.* one unit updates at a time), zero threshold (iff weighted inputs are ≥ 0 the unit is active) and outputs in {–1,1}. Draw this as a graph in which the possible states are nodes and the possible transitions are directed edges.
5. Which of the states are stable? (tip: there are two stable states)