Proposal

We seek to investigate from both an experimental and theoretical perspective the convergence properties of the Hebbian learning model. In this model, we consider a bidirectional complete graph of $G = (V, E)$ consisting of $|V| = n$ vertices representing neurons and $|E| = n(n-1)$ edges representing synapses between these neurons.

We model the vertices of G as distributed autonomous agents in order to impose biological locality constraints. The graph updates in rounds, and in each round neurons fire with probability equal to the sum of neighbors who fired in the previous round. At round t, each weight $e = (u, v)$ is updated accorded to the Hebbian rule where the weight of e is increased by a factor of $1+\epsilon$ if u fires in round t and v fires in round $t+1$, after which all incoming weights into a vertex are normalized. In order to formalize this update rule, we introduce an edge weight function $w_t : E \to \mathbb{R}$ that gives the weight of edge e at round t, and a indicator function $f_t(v) = \mathbb{I}[\text{Vertex v fires at round t}]$ which is 1 if vertex v fires at round t and 0 otherwise. The update rule is then

$$w_{t+1}((u,v)) = (1 + f_t(u) f_{t+1}(v)\epsilon) w_t((u,v))$$

We note that in order to prevent blowup of edge weights we further normalize the weights at the end of round t such that

$$w_t((u,v)) \leftarrow \frac{w_t((u,v))}{\sum_{\{x|(x,v)\in E\}} w_t((x,v))}$$

We initialize our system so that $f_0(v)$ is set to a random value in $\{0, 1\}$ for all $v \in V$, and all edge weights are set to a random value in $[0, 1]$.

We wish to study several properties of this system as it evolves. First, we seek to investigate experimentally if this system converges to fixed values as $t \to \infty$. If this seems to be the case, we will attempt to provide a theoretical justification for when and why this convergence will occur.

We next seek to tie this model to properties of biological relevance. In particular, establishing experimental and possibly theoretical bounds on the size of connected components in the graph may have connections to how memories are stored in the brain.

We finally seek to establish how this model can be used for learning. In order to do so, we will modify the graph such that certain nodes are fired based on external inputs rather than the probabilistic model, and we will observe how different types of input propagate through the system. Based on biological intuition, different regions of the graph should be responsible for processing different classes of inputs, and we hope to see the emergence of such properties in our model. Another extension to the learning aspect of this model would be to have another set of neurons act as a "output" and to use the correctness of this output as compared to a classification of the input to cause the graph G to evolve in such a way that classifications tend to be correct - this may further allow us to investigate the pathways by which patterns are learned.

In all, we hope to produce several experimental results that allow us to better understand the graph-theoretic properties of the Hebbian model and to relate these properties to biological learning. A further goal is to give theoretical justification for these experimental results if we are able to make headway on these problems.

References

[1] Nancy Lynch, Cameron Musco, Merav Parter. Computational Tradeoffs in Biological Neural Networks: Self-Stabilizing Winner-Take-All Networks, ITCS17.

[2] Nancy Lynch, Cameron Musco, Merav Parter. Neuro-RAM unit with Applications to Similarity Testing and Compression in Spiking Sigmoidal Networks, DISC17

[3] Robert Legenstein, Wolfgang Maass, Christos Papadimitriou, Santosh Vempala. Long Term Memory and the Densest K-Subgraph Problem, ITCS18

[4] Gerstner, Wulfram. Hebbian learning and plasticity. From neuron to cognition via computational neuroscience (2011): 0-25.