Department of Statistics

DISSERTATION PROPOSAL

ULISES PEREIRA

Department of Statistics
The University of Chicago

Attractor Neural Networks Models Consistent with Data

WEDNESDAY, November 30, 2016, at 2:00 PM
Jones 304, 5747 S. Ellis Avenue

ABSTRACT

The attractor neural network (ANN) scenario has been a popular scenario for memory storage in the association cortex, but there is still a large gap between these models and experimental data. In this talk we discuss three particular issues that have surprisingly received little attention: whether the distribution of the learned patterns is compatible with data; whether the learning rules used in such models are compatible with data; whether the temporal evolution of single neuron activity is compatible with data. A recent study (Lim et al, 2015) has found in IT cortex a distribution of neuronal responses close to lognormal, at odds with bimodal distributions of firing rates used in the vast majority of theoretical studies; and a Hebbian learning rule dominated by depression with a non-linear dependence on postsynaptic firing rate. On the other hand, multiple studies have shown high temporal variability in single units during delay periods in delay match to sample (DMS) experiments. This is at odds with the classical ANN's view of the delay period, where the network state converges to a static attractor correlated with one of the memories but where activities of neurons do not vary in time.

We study an ANN model in which external inputs defining the stored patterns have a unimodal distribution, with a family of generalized Hebbian rules that captures the recent findings in Lim et al, 2015. Using a mean field approach, we show that the learning rule that optimizes storage capacity has: (1) a highly non-linear dependence on the pre and postsynaptic firing rate; (2) a bias towards depression of the post-synaptic non-linearity; (3) A threshold between depression and potentiation that is much higher than the mean firing rate. All these features are consistent with the learning rules recently inferred from data. We derive distributions of firing rates for novel and familiar stimuli during both presentation and delay period, finding unimodal distributions consistent with data from DMS experiments. These results suggest that learning rules inferred from data in IT cortex are optimal for memory storage. In the later part of the talk, we show that this model presents a chaotic phase with associative memory properties. We argue that this can be a mechanism that reconciles ANN models with the single unit temporal variability observed during delay periods.
