Open to undergrads and grads

9.53: Emergent Computation within Distributed Neural Circuits

Brains and computers process information in radically different ways, with the constitutive elements of brains – synapses and neurons – exhibiting decidedly inferior performance characteristics, in terms of transmission speed, clock rate, signal-to-noise ratio, etc. Yet somehow, brains still outperform the best computer algorithms in most domains of sensory, motor, and cognitive function. Here we explore the emergent computational mechanisms and principles by which neural ensembles collectively instantiate remarkable behavioral competencies, despite the inherent limitations of messy biological wetware. This biologically-grounded perspective provides important insights in the domains of both neurobiology and neuro-inspired artificial intelligence. On the neurobiological side, we look to identify signatures of these computational mechanisms in actual neurophysiological data; on the artificial intelligence side, we gain an understanding of which biological motifs have been algorithmically implemented and which have not.

Neural Network Basics

- The McCulloch-Pitts neuron, classification task vs. functional mapping
- Levels of abstraction: binary threshold unit, firing rate approximation w. linear sum + sigmoidal non-linearity, leaky integrate-and-fire, Hodgkin-Huxley.
- Simple Perceptron and learning convergence, LMS algorithm, Minsky-Papert critique.
- Different types of learning: unsupervised, supervised, reinforcement.
- Basic back-propagation (Werbos, Rumelhart et al.), kernel-based methods (radial basis functions) function approximation capability, gradient descent, generalization, biological plausibility.

Dynamical Systems and Memory Models

- Stability, types of equilibrium points, attractor networks, Lyapunov functions.
- Neurodynamical models and content-addressable memory: discrete Hopfield model, pattern completion, storage capacity, energy function, continuous Hopfield model, Cohen-Grossberg Theorem.
- Associative learning vs. auto-associative memory.
- Other models of memory: expansion recoding (Marr/Albus model of the cerebellum), Kanerva's sparse distributed memory, Marr model of the hippocampus.
- Experimental data on memory: evidence of the cerebellum as an expansion recoder, CA3 as an attractor network, system-level consolidation, synaptic trace theory, engram cells.

Principles of Self-Organization

- Unsupervised learning basics: Hebbian learning, PCA, Oja's rule, Sanger's rule.
- 1-D and 2-D self-organizing feature maps (Kohonen), models of topography in cortex, Von der Malsburg orientation tuning model,
- Competitive learning: recurrent competitive fields as automatic "gain control" (Grossberg), winner-take-all networks.
- Reinforcement learning, differences with supervised learning, level of biological plausibility.
- Experimental data: feature maps in the visual/auditory systems, representation of acoustic sensorimotor structure in birdsong from beginner to expert, evidence of "gain control in V1.

Feedback, Recurrence, and the Stability-Plasticity Dilemma

- Back-propagation through time (Werbos), recurrent neural networks, echo state networks and reservoir computing, Trade-off between Interference and generalization in distributed circuits, how the problem is solved in machine learning, biological solutions, synaptic dynamics.
- Adaptive resonance theory (Carpenter and Grossberg).
- Experimental data: behavioral results on interleaved practice/generalization, practice-dependent response properties of single neurons. 2-photon microscopy studies on dendritic spine turnover during learning,

Noise in Neural Circuits, Representation, and Decoding

- Noise levels in real neural circuits, models of noisy processing, hyperplastic and noisy networks.
- Decoding signals in a complex system, causality vs. correlation, the concept of representation, neuroprosthetic devices.
- Experimental data: V1, M1, and A1 response properties, the efficacy of decoding algorithms in brain-machine interfaces, ways forward.