A mathematical theory of semantic development

Andrew Saxe, James L. McClelland, and Surya Ganguli

Many psychology experiments have revealed remarkable regularities in the developmental time course of semantic cognition in infancy and childhood, as well as its progressive disintegration in adult dementia. For example, infants tend to learn to make broad categorical distinctions between concepts (i.e. plant/animal) before they can make finer scale distinctions (i.e. dog/cat), and this process is reversed in dementia, where finer scale distinctions are lost before broad ones. What are the theoretical principles underlying the ability of neuronal networks to discover categorical structure from experience?

We address this question by developing a phenomenological, mathematical theory of semantic development through an analysis of the learning dynamics of multilayer networks exposed to hierarchically structured data. We find new exact solutions to the nonlinear dynamics of error corrective learning in deep, 3 layer networks. These solutions reveal that networks learn input-output covariation structure on a time scale that is inversely proportional to its statistical strength. We further analyze the covariance structure of data sampled from hierarchical generative models, and show how such models yield a hierarchy of input-output modes of differing statistical strength, leading to a hierarchy of time-scales over which such modes are learned.

Our results reveal, quite strikingly, that even the second order statistics of hierarchically structured data contain powerful statistical signals sufficient to drive many complex experimentally observed phenomena in semantic development, including the progressive differentiation of concepts and its reversal in dementia, sudden stage-like transitions in performance, erroneous illusory correlations in early learning, and quantitative theories relating category coherence to learning speed. Moreover, our work reveals how deep network structure is essential for exhibiting these phenomena.

Beyond semantic cognition, our analytical results provide an extremely general formalism for understanding how the statistical structure of experience drives learning in deep networks, and could provide insight into the learning dynamics of many different network models in psychology and neuroscience.