Abstract Submitted
for the MAR17 Meeting of
The American Physical Society

Feature to prototype transition in neural networks DMITRY KROTOV[1], Institute for Advanced Study, JOHN HOPFIELD[2], Princeton University — Models of associative memory with higher order (higher than quadratic) interactions, and their relationship to neural networks used in deep learning are discussed. Associative memory is conventionally described by recurrent neural networks with dynamical convergence to stable points. Deep learning typically uses feedforward neural nets without dynamics. However, a simple duality relates these two different views when applied to problems of pattern classification. From the perspective of associative memory such models deserve attention because they make it possible to store a much larger number of memories, compared to the quadratic case. In the dual description, these models correspond to feedforward neural networks with one hidden layer and unusual activation functions transmitting the activities of the visible neurons to the hidden layer. These activation functions are rectified polynomials of a higher degree rather than the rectified linear functions used in deep learning. The network learns representations of the data in terms of features for rectified linear functions, but as the power in the activation function is increased there is a gradual shift to a prototype-based representation, the two extreme regimes of pattern recognition known in cognitive psychology.

[1]Simons Center for Systems Biology
[2]Princeton Neuroscience Institute

Dmitry Krotov
Institute for Advanced Study

Date submitted: 28 Oct 2016 Electronic form version 1.4