Interacting perceptrons: a minimal model of spontaneous collective learning

Jérôme Garnier-Brun[1], Sebastiano Ariosto[1], Davide Straziota[1] and Luca Saglietti[1]

[1]Department of Computing Sciences, Bocconi University, Milano

We introduce a minimal model of unsupervised learning in which linear classifiers update their parameters by enforcing consistency with their own predictions on a stream of fresh data, regularized by a parsimony prior. Despite the absence of ground-truth labels or rewards, individual agents can spontaneously align with the latent structure of the data-generating environment, stabilizing into self-sustaining modes of operation we term functional replicators. Using tools from statistical mechanics, we derive the phase boundary separating learning from non-learning regimes for a single agent. We then extend the framework to a population of agents that pool labels from their peers, deriving how the phase boundary shifts with the number of labeling agents and identifying regimes where collective interaction either facilitates or hinders learning. Numerical simulations reveal a crossover in the collective dynamics from voter-like, fluctuation-driven consensus to drift-driven mean-field alignment as the pool size grows. Our results suggest a minimal setting for studying decentralized learning as a self-organizing dynamical process.