Learning and generalizing sameness in simple recurrent networks

Juan Valle-Lisboa

Secccion Biofisica, Facultad de Ciencias, UDELAR, URUGUAY

Abstract: I study the ability of simple recurrent networks to learn and generalize sameness. To this end, I trained 100 networks to output 1 if two successive real valued vectors in a sequence were similar and 0 otherwise. Although learning of the training set proceeded easily, we could not make any of the networks generalize beyond the training set, despite the test vectors being "within the training space". In an attempt to understand these limitations I studied networks of threshold units. In the first place I analyzed networks able to process sequences of bits. I show that part of the problems arise because, depending on the training set, there might be cases were not every accessible internal state of the network is visited, leaving those cases "out of the training space". The case of multidimensional inputs is even more difficult given that the networks get trapped in wrong generalizations.