

The statistical physics of learning in under- and over-parameterized layered neural networks

Otavio Citton^{1,2}, Frederieke Richert¹, Michael Biehl¹

¹University Of Groningen, Groningen, The Netherlands, ²Groningen Cognitive Systems and Materials Center (CogniGron), Groningen, The Netherlands

Using techniques borrowed from statistical physics, we study neural network systems at thermal equilibrium in a student-teacher model setting. The architecture studied here is the so-called Soft Committee Machine, a simple two-layered feedforward neural network with square integrable activation function. Our focus is on understanding the effects that a mismatch in the number of internal units between the student and the teacher networks has in the system's behavior. We introduced two new ansätze for the saddle points of the under- and over-parameterized systems (supported by Monte Carlo simulations) and obtain the corresponding free energy and state equations for systems with any degree of over-parameterization and any square integrable activation in the high temperature limit. We show that the critical point grows with the degree of over-parameterization, while the asymptotic error decreases up until the number of internal units is the same as the teacher network's. For larger degrees of over-parameterizations the system always converges to a perfect generalization state. Moreover, we also show that, in an over-parameterized setting, out of all the activation functions analyzed here, the ReLU is the one with the highest decay rate for the order parameter that controls the baseline error created by the extra units.

