Artificial neural networks are a class of models used in machine learning , and inspired by biological neural networks . They are the core component of modern deep learning algorithms. Computation in artificial neural networks is usually organized into sequential layers of artificial neurons . The number of neurons in a layer is called the layer width. Theoretical analysis of artificial neural networks sometimes considers the limiting case that layer width becomes large or infinite. This limit enables simple analytic statements to be made about neural network predictions, training dynamics, generalization, and loss surfaces. This wide layer limit is also of practical interest, since finite width neural networks often perform strictly better as layer width is increased. [1] [2] [3] [4] [5] [6]

Theoretical approaches based on a large width limit

The Neural Network Gaussian Process (NNGP) corresponds to the infinite width limit of Bayesian neural networks, and to the distribution over functions realized by non-Bayesian neural networks after random initialization. [7] [8] [9] [10]

The same underlying computations that are used to derive the NNGP kernel are also used in deep information propagation to characterize the propagation of information about gradients and inputs through a deep network. [11] This characterization is used to predict how model trainability depends on architecture and initializations hyper-parameters.

The Neural Tangent Kernel describes the evolution of neural network predictions during gradient descent training. In the infinite width limit the NTK usually becomes constant, often allowing closed form expressions for the function computed by a wide neural network throughout gradient descent training. [12] The training dynamics essentially become linearized. [13]

Mean-field limit analysis, when applied to neural networks with weight scaling of $\sim 1/h$ instead of $\sim 1/\sqrt{h}$ and large enough learning rates, predicts qualitatively distinct nonlinear training dynamics compared to the static linear behavior described by the fixed neural tangent kernel, suggesting alternative pathways for understanding infinite-width networks. [14] [15]

Catapult dynamics describe neural network training dynamics in the case that logits diverge to infinity as the layer width is taken to infinity, and describe qualitative properties of early training dynamics. [16]

References