

## Pietro Rotondo: Statistical mechanics of deep learning beyond the infinite-width limit

Decades-long literature testifies to the success of statistical mechanics at clarifying fundamental aspects of deep learning. Yet the ultimate goal remains elusive: we lack a complete theoretical framework to predict practically relevant scores, such as the train and test accuracy, from knowledge of the training data. Huge simplifications arise in the infinite-width limit, where the number of units  $N_\ell$  in each hidden layer ( $\ell=1, \dots, L$ , being  $L$  the finite depth of the network) far exceeds the number  $P$  of training examples. This idealisation, however, blatantly departs from the reality of deep learning practice, where training sets are larger than the widths of the networks. Here, we show one way to overcome these limitations. The partition function for fully-connected architectures, which encodes information about the trained models, can be evaluated analytically with the toolset of statistical mechanics. The computation holds in the "thermodynamic limit" where both  $N_\ell$  and  $P$  are large and their ratio  $\alpha_\ell = P/N_\ell$ , which vanishes in the infinite-width limit, is now finite and generic. This advance allows us to obtain (i) a closed formula for the generalisation error associated to a regression task in a one-hidden layer network with finite  $\alpha_1$ ; (ii) an approximate expression of the partition function for deep architectures (technically, via an "effective action" that depends on a finite number of "order parameters"); (iii) a link between deep neural networks in the proportional asymptotic limit and Student's  $t$  processes; (iv) a simple criterion to predict whether finite-width networks (with ReLU activation) achieve better test accuracy than infinite-width ones. As exemplified by these results, our theory provides a starting point to tackle the problem of generalisation in realistic regimes of deep learning.