Institut for Theoretical Physics
University Leipzig

Prof. Dr. B. Rosenow
A. Afanah , M. Staats

Advanced Statistical Physics - Sheet 1

Summer Terms 2022

Hand in: Hand in tasks marked with * to mailbox no. (43) inside ITP room 105b by Friday 22.04. at 9:15 am.

3. Correlation function for an ultrawide network *12* Points*

Here you are asked to compute the correlation function of the network outputs for a network with one hidden layer. The calculation is performed by averaging over all possible weights of the network. This result applies to an infinitely wide network which is randomly initialized, such that the correlation function is self-averaging, i.e. the correlation function of a given network is equivalent to the explicit average over the network weights. The correlation function is also called kernel function and can be used to approximate the network prediction after training.

We consider a network with one hidden layer of width N, input dimension d, and output function

$$f_{\mathbf{w}}(\mathbf{x}) \;=\; \sum_{i=1}^{N} w_i^{(2)} \phi(\mathbf{w}_i^{(1)} \cdot \mathbf{x}) \quad .$$

Here, both the input vector $\mathbf{x}$ and the vectors $\mathbf{w}_i^{(1)}$ are d-dimensional. The correlation function (also called kernel function) is defined as an average over all weights

$$k(\mathbf{x}, \mathbf{y}) \;=\; \langle f_{\mathbf{w}}(\mathbf{x}) f_{\mathbf{w}}(\mathbf{y}) \rangle_{\mathbf{w}} \quad .$$

Compute the kernel function for a case in which the elements of the $\mathbf{w}_i^{(1)}$ are i.i.d. Gaussian distributed with variance σ_1^2 and mean zero, the $w_j^{(2)}$ are Gaussian i.i.d with variance σ_2^2 and mean zero, and the activation function is ReLu, i.e. $\phi(z) = z\theta(z)$.

Hint: it may be useful to consult the publication

Y. Cho and L. K. Saul, *Kernel methods for deep learning*, in Proceedings of the 22nd International Conference on Neural Information Processing Systems, NIPS'09 (Curran Associates, Red Hook, NY , 2009), pp. 342-350.