Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

A thermodynamic approach to deep learning

Relatore:
Prof.ssa Rita Fioresi

Correlatore:
Prof. Francesco Faglioni

Presentata da:
Francesco Morri

Anno Accademico 2019/2020

Abstract

Neural Networks are an incredibly powerful tool used to solve complex problems.
The actual functioning of this tool and its behaviour when applied to different kind of problems is not completely explain though.
In this work we study the behaviour of a neural network, used to classify images, through a physical model, based on statistical thermodynamics. We found interesting results regarding the *temperature* of the different components of the network, that may be exploited in a more efficient training algorithm.

Contents

Chapter 1

Neural Networks

In this chapter we study neural networks. Starting from a brief historical digression, we then introduce a non trivial example of neural network, the linear classifier, which is an evolution of the first ever attempted approach: the *perceptron*. We eventually describe more complex networks used in image recognition problems, the *convolutional neural networks*.
We are also going to explain the process of training and optimizing a neural network, using as toy example the linear classifier, which is easier to understand and describe in its functioning, with respect to convolutional neural networks.

Chapter 2

Dynamics of the Model

In this chapter we introduce the physical model we want to use to describe a convolutional neural network. Our formulation of the model employs the theory of statistical thermodynamics: we first start by explaining the parallelism between a neural network and a thermodynamical system, then, we present our interpretation of the actual dynamics of the network.
We conclude the chapter with a section about how the thermodynamical concept of temperature can be expressed in the system of the neural network, a topic we investigate further with numerical simulations.

Chapter 3

Simulation Results

In this chapter we introduce the architecture used and we report the results of the numerical simulation.
Inspired by the work of P. Chaudhari and S. Soatto ([14]), that linked the concept of temperature of the network to the learning rate η and batch size $|\mathbb{B}|$, we focused on finding the relation between these quantities, that is $T(\eta, |\mathbb{B}|)$. In order to achieve this, we changed the two values of η and β to observe how the velocities, and the temperatures, of the layers varied.
We also show the differences that occur inside a single layer, between all the weights.

Bibliography

[1] Y. LeCun; C. Cortes; C. J. C. Burges. *MNIST DATABASE*. `http://yann.lecun.com/exdb/mnist/`.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. *ImageNet: A Large-Scale Hierarchical Image Database*. In *CVPR09*, 2009.

[3] W. S. McCulloch; W. Pitts. *A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics*, 5, 12 1943.

[4] F. Rosenblatt. *The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain. Psychological Review*, pages 65–386, 1958.

[5] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. *Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation*, 1:541–551, 1989.

[6] K. Fukushima. *Neocognitron. Scholarpedia*, 2(1):1717, 2007. revision #91558.

[7] D. Hubel and T. Wiesel. *Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Journal of Physiology*, 160:106–154, 1962.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*, 2015.

[10] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. *Striving for Simplicity: The All Convolutional Net*, 2014.

[11] Sergey Ioffe and Christian Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, 2015.

[12] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R. Salakhutdinov. *Improving neural networks by preventing co-adaptation of feature detectors*, 2012.

[13] Yann Lecun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. *Gradient-based learning applied to document recognition.* In *Proceedings of the IEEE*, pages 2278–2324, 1998.

[14] P. Chaudhari; S. Soatto. *Stochastic gradient descent performs variational inference, converges to limit cycles for deep networks*, 2017.

[15] I. Goodfellow; Y. Bengio; A. Courville. *Deep Learning.* MIT Press, 2016. `http://www.deeplearningbook.org`.

[16] J. Heaton. *AIFH, Volume 3: Deep Learning and Neural Networks.* December 31, 2015, 2015.

[17] Y. LeCun; Y. Bengio; G. Hinton. *Deep learning. Nature*, 521, 5 2015.

[18] E. Fermi. *Termodinamica.* Programma di matematica, fisica, elettronica. Bollati Boringhieri, 2. ed., 6. rist edition, 1988.

[19] L. D. Landau; E. M. Lifsits; L. P. Pitaevskij. *Fisica Statistica, Parte Prima*, volume 5 of *Fisica Teorica.* Editori Riuniti, 1986.