Available at: http://www.ictp.trieste.it/~pub_off IC/99/130

United Nations Educational Scientific and Cultural Organization
and
International Atomic Energy Agency

THE ABDUS SALAM INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS

ON THE FINITE SIZE SCALING IN NEURAL NETWORKS

E. Korutcheva[1]
Dep. Fisica Matematica y Fluidos, UNED,
c/Senda del Rey No 9, 28080 Madrid, Spain,[2]
G. Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences,
1784 Sofia, Bulgaria
and
The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

and

N. Tonchev
G. Nadjakov Institute of Solid State Physics, Bulgarian Academy of Sciences,
1784 Sofia, Bulgaria.

Abstract

We study the first-order phase transition in the model of a simple perceptron with continuous weights and with large, but finite values of the inputs. Making the analogy with the first-order phase transitions in usual finite-size physical systems, we calculate the shift and the rounding exponents near the transition point. In the case of a general perceptron with larger variety of outputs, the analysis could only give bounds for the exponents.

MIRAMARE – TRIESTE
September 1999

---

[1]Regular Associate of the Abdus Salam ICTP.
[2]Current address. Fax: +34-91-398-66-97; E-mail: elka@fisfun.uned.es

Recently W. Nadler and W. Fink [1] showed that the fraction of patterns that can be stored in a single and a multilayer neural network (perceptron) obeys a finite-size scaling (FSS) behavior. A similar FSS result was previously demonstrated in some problems such as the satisfiability of random boolean expressions, the connectivity of random graphs [2] and the quasi-species model of molecular evolution [3], where no intrinsic length scale is present.

The aim of the present comment is to complement the FSS study of perceptron using some previous knowledge from FSS behavior in different physical systems [4] - [6].

The system, we are interested in, is a single layer perceptron storing a set of input patterns $\{\xi_i^\mu\}, i = 1, ..., N; \mu = 1, ..., p$, drawn from a Gaussian distribution. By N and p we denote the numbers of the inputs and the patterns, respectively.

It is well known that for this system, the fraction of all possible input-output relations of size $\alpha = p/N$ that can be stored, called $P(\alpha, N)$, exhibits a smooth transition from one to zero, which becomes discontinuous, i.e. first-order one, at the critical storage capacity α_c and in the thermodynamic limit $N \to \infty$ [7, 8]. In the case of a single- layer perceptron with one output unit, Gaussian inputs and continuous couplings, an exact result for $P(\alpha, N)$ is known [8]:

$$P(\alpha, N) = 2^{1-p} \Sigma_{i=0}^{N-1} \binom{p-1}{i}. \tag{1}$$

When the size of the system N is large, eq.(1) takes the asymptotic form:

$$P(\alpha, N) \approx \frac{1}{2} \left(1 + Erf\left(\sqrt{\frac{N}{2\alpha}}(2-\alpha)\right) \right), \tag{2}$$

revealing an FSS behavior with a scaling parameter

$$y = (\alpha - \alpha_c) N^{1/\nu} \tag{3}$$

and a scaling exponent $\nu = 2$ near the transition point $\alpha_c = 2$. Thus it is interesting to investigate the finite-size effects near the transition point and of particular interest is the determination of the shift and the rounding of the transition.

Following the analogy with the conventional first-order transitions [4]-[6], we define as a transition point $\alpha_c(N)$ this value of the parameter α, for which the derivative $\left|\frac{\partial P(\alpha,N)}{\partial \alpha}\right|$ shows a maximum for large but finite values of N (N being the size of the system). This derivative becomes divergent in the thermodynamic limit at $\alpha_c = 2$. The above definition permits to treat in a similar way the finite-size effects in a perceptron with continuous and binary weights.

From eq.(1) we calculated the inflection point of the function $P(\alpha, N)$ with respect to the parameter α for different values of N and we identified the so-called shift exponent [4]-[6]

$$\alpha_c(N) - \alpha_c(\infty) \sim \frac{1}{N^\lambda}. \tag{4}$$

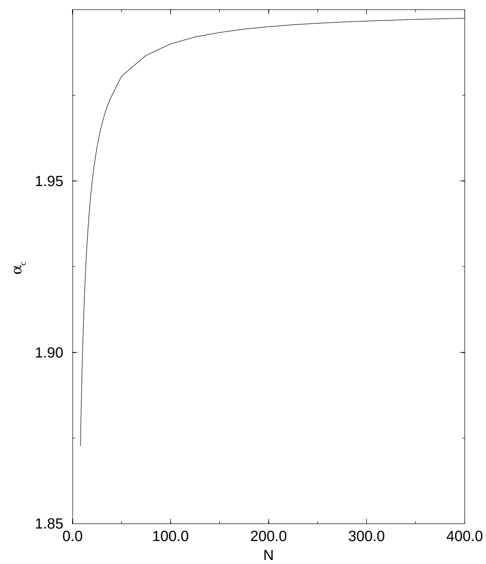

Figure 1: The critical storage capacity $\alpha_c(N)$ for values of N within the interval [8,400].

We did the same analysis also by using eq.(2). We obtained the following dependence of the critical storage capacity $\alpha_c(N)$ for N-large (see Fig.1):

$$\alpha_c(N) - \alpha_c(\infty) \sim \frac{1}{N}, \tag{5}$$

giving a shift exponent $\lambda = 1$.

The result (5) is similar to the well-known result for asymmetric temperature-driven first-order transitions in finite d-dimensional systems with cubic symmetry, where the location of the shift by the maximal slope scales like L^{-d} (L being the finite size) [4]- [6]. Note that in our case the system is effectively one-dimensional (of finite size N) and there are no boundary conditions imposed.

The two classes of transitions, symmetric and asymmetric, however show a rounding behavior, which is given by the scaling of the width of the peak of the diverging observable. In other words, it is the interval over which the singularity is smeared out and which becomes increasingly sharp as the finite dimension of the system goes to infinity.

In the concrete case of the simple perceptron this is the scaling of the width of $\left|\frac{\partial P(\alpha,N)|}{\partial\alpha}\right|$, which determines the rounding behavior. Using eq.(2):

$$\left|\frac{\partial P(\alpha, N)}{\partial\alpha}\right| \sim \sqrt{\frac{N}{2\alpha}}\exp\{-\frac{[N(2-\alpha)^2]}{2\alpha}\}, \tag{6}$$

the rounding exponent $\theta = \frac{1}{2}$ follows. Note that a similar behavior with N occurs for the shift and the variance of the generalization error in the case of a Bayesian perceptron with continuous weights [9]. Their specific dependence on α however cannot be explained by the present analysis.

An interesting problem is what happens in the case of a perceptron with binary weights [10]. For this case the numerical analysis for the typical fraction shows a sharpness between the two regimes by increasing N, but there is no definitive conclusion about the step-like behavior in the thermodynamic limit [11]. An investigation, similar to that for the continuous case, could explain many of the features of this behavior [12].

In the general case of a perceptron, having a larger variety of outputs [13], it has been proved, [14], that for $p \geq d_{VC}$ (d_{VC} being the Vapnik-Chervonenkis (VC) dimension), the fraction of all possible input-output relations obeys the following inequality:

$$P(\alpha, N) \leq 2^{1-p} \Sigma_{i=0}^{d_{VC}} \left(\begin{array}{c} p-1 \\ i \end{array} \right). \tag{7}$$

It has been shown [15] that in the thermodynamic limit $N \to \infty$ $(p, d_{VC} \to \infty)$ and keeping $\alpha = \frac{p}{N}$, and $\alpha_{VC} = \frac{d_{VC}}{N}$ fixed, the VC-entropy shows different behavior above and below $\alpha = 2\alpha_{VC}$, which permits to relate the storage capacity of the network to its VC- dimension via $\alpha_c \leq 2\alpha_{VC}$ $(\alpha_c \equiv \alpha_c(\infty))$.

Eq.(7) shows that for N-large, the asymptotic form of the upper bound $\bar{P}(\alpha, N)$ of the fraction $P(\alpha, N)$ is given by:

$$P(\alpha, N) \leq \bar{P}(\alpha, N) = \frac{1}{2} \left[1 + Erf \left(\sqrt{\frac{N}{2\alpha}} (2\alpha_{VC} - \alpha) \right) \right], \tag{8}$$

leading to the same values for the shift and the rounding exponents for the upper bound, known from the case of the simple perceptron, These are also upper bounds for the shift and rounding exponents in the general case. Note also, that for this case, the VC analysis gives only bounds for the d_{VC} [11], [15].

Finally, we would like to mention that considering the behavior of the VC-entropy above and below the transition, or studying the behavior of the annealed entropy and its relation to the generalization error, leads again to the problem of finding inequalities for the functions of interest [15].

Obviously, the full understanding of the general problem requires an investigation for every concrete case of architecture and machine [12].

Acknowledgments: We thank the Abdus Salam International Centre for Theoretical Physics, Trieste, Italy, for hospitality and support and where this work were completed. This work was done within the framework of the Associateship Scheme of the Abdus Salam ICTP. E.K. thanks M. Gordon, W. Kinzel, M. Opper and A.Engel for discussions, and the Workshop on "Statistical Physics of Neural Networks", organized by the Max-Planck Institute-Dresden, 1999, where parts of this work have been presented. This work is also supported by the Spanish DGES Project PB97-0076 and the Bulgarian Scientific Foundation Grant F-608.

References

[1] W.Nadler and W.Fink, Phys.Rev.Lett. **78**(1997)555.

[2] S.Kirkpatrick and B.Selman, Science **264**(1994)1297.

[3] P.R.A.Campos and J.F.Fontanari, Phys.Rev. **E58**(1998) 2664.

[4] V.Privman, Ed., *Finite-size Scaling and Numerical Simulations of Statistical Systems*, World Scientific, Singapore, 1990.

[5] M.Barber in:*Phase Transitions and Critical Phenomena*, Eds. C.Domb and J.L.Lebowitz, Academic, London, 1983, vol.8.

[6] K.Binder, in *The Monte Carlo Method in Condensed Matter Physics*, Springer-Verlag, 1992.

[7] J.Hertz, A.Krogh and R.G.Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley,Reading, MA, 1991.

[8] T.M.Cover, IEEE Trans.Electron.Comput.**14**(1965)326.

[9] A.Buhot, J.-M.Torres Moreno and M.Gordon, Phys.Rev.**E55** (1997)7434.

[10] W.Krauth and M.Mézard, J.Phys.(Paris) **50** (1990) 2613.

[11] S.Mertens and A.Engel, Phys.Rev.**E55** (1997)4478.

[12] E.Korutcheva, in progress.

[13] E.Barkai, D.Hansel and H.Sompolinsky, Phys.Rev. **A45**(1992) 4146.

[14] V.Vapnik, *Estimation of Dependencies Based on Empirical Data*, Springer-Verlag, 1982.

[15] M.Opper, Phys.Rev **E51** (1995) 3613.