On the computational capability of a perceptron

1993 J. Phys. A: Math. Gen. 26 L1233

(http://iopscience.iop.org/0305-4470/26/23/008)

J Phys. A: Math. Gen. 26 (1993) L1233–L1238. Printed in the UK

LETTER TO THE EDITOR

On the computational capability of a perceptron

J F Fontanari† and W K Theumann ‡

† Instituto de Física e Química de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560 São Carlos SP, Brazil
‡ Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixo Postal 15051, 91501 Porto Alegre RS, Brazil

Received 30 July 1993

Abstract. The minimal fraction of errors a perceptron can make on P random dichotomies is investigated. An exact upper bound is presented as well as comparisons between the replica theory estimates and a rigorous lower bound.

The computational capability of a perceptron, also termed formal neuron or linear threshold element, is an issue of interest in such diverse research fields as machine learning and statistical mechanics. In fact, the perceptron is the basic computing device in many complex neural networks, hence the interest to understand its computational limitations. Moreover, it is the only continuous, infinite range spin glass model that possesses a phase where the replica symmetry is broken. Unfortunately important contributions from each of these fields are usually not properly appreciated by the others. The main goal of this note is to compare a rigorous lower bound due to Venkatesh and Psaltis (1992) for the minimal fraction of errors ε that a perceptron makes on P binary decisions with the average case results of Gardner's statistical mechanics framework (Gardner 1988). We derive also an exact upper bound for ε.

The perceptron we consider in this note is a device which accepts N binary (± 1) inputs $\boldsymbol{S} = (S_1, S_2, \ldots, S_N)$ and produces a single output bit σ given by the sign of a weighted sum of its inputs,

$$\sigma = \mathrm{sign}\left(\sum_{i=1}^{N} W_i S_i \right). \tag{1}$$

The computational capability of such a device is measured by the maximum number of random decisions $S^l \to \sigma^l; l = 1, \ldots, P$ it can reliably make. It was shown by Cover (1965) that, in the limit $N \to \infty$, the maximum number of random decisions a perceptron can realize without making errors is $P = 2N$. This result was later re-derived by Gardner (1988) within the statistical mechanics framework.

A related issue addressed in this note is how this computational capability is improved if one allows the perceptron the freedom to make a fixed number of errors on the P decisions. In this line, Venkatesh and Psaltis (1992) have demonstrated that, in the limit $N \to \infty$, there is no choice of weights $\boldsymbol{W}$ for which a perceptron makes fewer than εP errors if the training set size is smaller than

$$P = \alpha N = \frac{2K_\varepsilon}{1 - 2\varepsilon} N. \tag{2}$$

0305-4470/93/231233+06$07.50

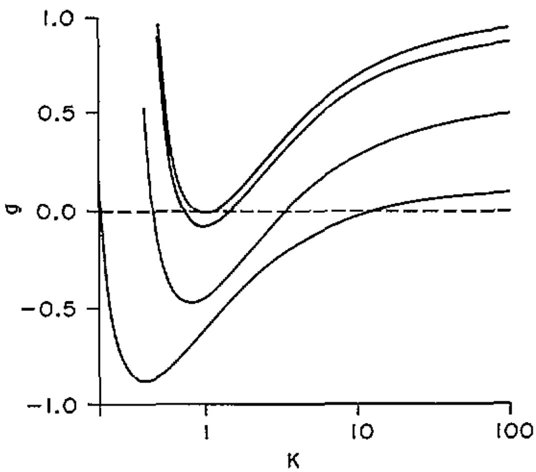

Figure 1. Function $g(K)$ for, from top to bottom, $\varepsilon=0.001$, 0.01, 0.1 and 0.3.

Here, K_ε is defined by the *unique* solution of the equation $g(K)=0$, where

$$g(K)\equiv 1-S\left(\frac{1-2\varepsilon}{2K}\right)-S(\varepsilon) \tag{3}$$

and $S(x)$ is the binary entropy function,

$$S(x)=-\frac{1}{\ln 2}\,(x\ln x+(1-x)\ln(1-x)). \tag{4}$$

Clearly, equation (2) gives a lower bound to the maximal storage capacity of a perceptron when a fixed fraction of errors ε is allowed. Moreover, it shows that this maximum number of decisions increases linearly with N, as in the case where no errors are allowed. Interestingly, the scale $P=\alpha N$ appears naturally in the statistical mechanics framework as the condition to obtain a well-defined thermodynamic limit. Before we proceed presenting the statistical mechanics framework, some remarks about the definition of K_ε are in order. In figure 1, the function $g(K)$ is depicted for several values of ε. This function possesses two roots so that K_ε is not uniquely defined by the equation $g(K)=0$ and therefore one must specify which root must be chosen so that equation (2) holds true. In fact, a condition for the validity of equation (2) is that $g(K_\varepsilon(1+\lambda))$ be positive for fixed but arbitrary $\lambda>0$ (Venkatesh and Psaltis 1992). As seen in figure 1, this condition is fulfilled by the largest root only. For $\varepsilon\to 0$ this root is given by $K\varepsilon\approx 1+\sqrt{-2\varepsilon\ln\varepsilon}$ so that $\alpha\approx 2+\sqrt{-8\varepsilon\ln\varepsilon}$, in disagreement with the prediction $\alpha\approx 2+4\varepsilon$ of Venkatesh and Psaltis (1992). As expected, for $\varepsilon=0$ equation (2) reduces to Cover's result (Cover 1965). Moreover, as pointed out by Brunel *et al* (1992), K_ε diverges for $\varepsilon\to\frac{1}{2}(\alpha\to\infty)$. Moreover specifically, for large α this lower bound is given by

$$\varepsilon_{VP}\approx\frac{1}{2}-0.76\sqrt{\frac{\ln\alpha}{\alpha}}. \tag{5}$$

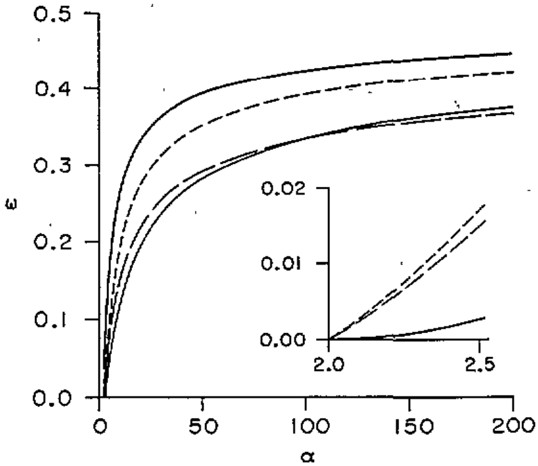

Figure 2. Several estimates for ε as function of α: rigorous lower bound (lower solid curve), replica-symmetric ansatz (long broken curve), one-step replica symmetry breaking ansatz (short broken curve) and exact upper bound (upper solid curve).

Fortunately, these details about the specific dependence of K_ε on ε are not relevant to the proof of equation (2). This lower bound is presented in figure 2 (lower solid curve) with α taken as the independent parameter.

The statistical mechanics approach to the problem of determining the computational capability of a perceptron consists basically in searching for the ground state of the energy function

$$E(W)=\sum_{l=1}^{P}\Theta\left(-\sigma^l\sum_{i=1}^{N}W_iS_i^l\right) \tag{6}$$

which counts the number of errors a perceptron with weights W makes on the P decisions $S^l\to\sigma^l$. Here, $\Theta(x)=1$ if $x>0$ and 0 otherwise. The minimal fraction of errors ε is then given by

$$\varepsilon=\lim_{\beta\to\infty}\frac{1}{\alpha}\frac{\partial\beta f}{\partial\beta} \tag{7}$$

where f is the average free-energy density

$$f=-\frac{1}{\beta N}\langle\ln Z\rangle \tag{8}$$

and Z is the partition function

$$Z=\int\prod_i \mathrm{d}W_i\,\delta\left(\sum_{i=1}^{N}W_i^2-N\right)\mathrm{e}^{-\beta E(W)}. \tag{9}$$

The parameter $\beta\equiv 1/T$ plays the role of the inverse temperature and models a fast noise acting on the weights. The symbol $\langle\ldots\rangle$ in equation (8) stands for the average over the distribution of the input patterns S^l and their respective output bits σ^l. Each component of each pattern S_i^l, as well as the output bits σ^l, are chosen randomly equal to $+1$ or -1 with the same probability. The spherical constraint enforced by the delta function in equation (9) is necessary to assure the integrals over W_i are finite. The average is performed through the replica method which consists in using the identity.

$$\langle \ln Z \rangle = \lim_{n\to 0} \frac{1}{n} \ln \langle Z^n \rangle \tag{10}$$

evaluating $\langle Z^n \rangle$ for *integer n* and then analytically continuing to real $n \approx 0$. According to Gardner and Derrida (1988), the final result for the average free-energy density is

$$f = -\lim_{n\to 0} \frac{1}{\beta n} \operatorname{extr} \left[\alpha G_0(q_{ab}) + G_1(\phi_{ab}, \eta_a) + \mathrm{i} \sum_{a<b} \phi_{ab}\, q_{ab} \right] \tag{11}$$

where

$$G_0(q_{ab}) = \ln \prod_{a=1}^{n} \int_{-\infty}^{\infty} \frac{\mathrm{d}\lambda_a}{2\pi} \mathrm{e}^{-\beta\Theta(-\lambda_a)} \int_{-\infty}^{\infty} \mathrm{d}x_a \exp\left[\mathrm{i} \sum_a x_a \lambda_a - \frac{1}{2} \sum_a x_a^2 - \sum_{a<b} q_{ab} x_a x_b \right] \tag{12}$$

$$G_1(\phi_{ab}, \eta_a) = \ln \int_{-\infty}^{\infty} \prod_{a=1}^{n} \mathrm{d}W_a \exp\left[\mathrm{i} \sum_a \eta_a (W_a^2 - 1) - \mathrm{i} \sum_{a<b} \phi_{ab} W_a W_b \right]. \tag{13}$$

The extremum is taken over the saddle-point parameters (η_a, q_{ab}, ϕ_{ab}), where η_a and ϕ_{ab} are Lagrange multipliers needed to enforce the spherical constraint and the definition of the physical order parameter

$$q_{ab} = \frac{1}{N} \sum_{i=1}^{N} W_i^a W_i^b \qquad a<b \tag{14}$$

respectively. The usual procedure adopted for evaluating the extremum in equation (11) is to assume an *ansatz* for the saddle point parameters η_a, q_{ab} and ϕ_{ab} so that the integrals in the expressions for G_0 and G_1 can explicitly be carried out. In this note we will consider two ansatz, namely, the replica symmetric (Sherrington and Kirkpatrick 1975) and the first step of Parisi's replica symmetry breaking scheme (Parisi 1980). For continuous weights perceptrons, the sole condition for the validity of these ansatz is their stability with respect to transverse fluctuations in the saddle-point parameters q_{ab} and ϕ_{ab} (de Almeida and Thouless 1978).

In the replica symmetric ansatz, one assumes that all order parameters are replica independent, i.e. $q_{ab} = q$, $\phi_{ab} = \phi$ and $\eta_a = \eta$. The region of validity of this solution in the plane (α, T), obtained by evaluating numerically the stability condition (de Almeida–Thouless line) found by Gardner and Derrida (1988), is shown in figure 3. For $\alpha > 2$ the replica symmetric saddle-point is unstable at low temperatures so that it could not be used to calculate the minimal fraction of errors, equation (7). For comparison purposes, however, we present ε calculated with this ansatz in figure 2 (long broken curve). Note that it violates the lower bound of Venkatesh and Psaltis for large α. In fact, in this regime we find

$$\varepsilon_{RS} \approx \frac{1}{2} - \left(\frac{3}{2\pi\alpha} \right)^{1/3} \tag{15}$$

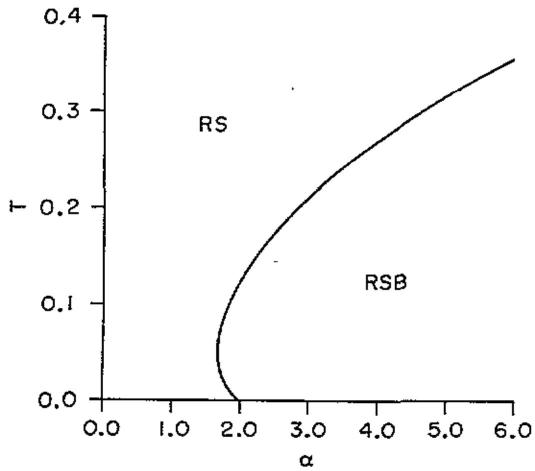

Figure 3. Phase diagram in the (α, T) plane, showing the de Almeida–Thouless line which delimits the region of validity of the replica-symmetric solution.

which should be compared with equation (5). Nevertheless, the fraction of errors at the de Almeida–Thouless line $\varepsilon_{\rm AT}$ can be calculated exactly with the replica–symmetric ansatz, giving an upper bound for ε since it is calculated at non-zero temperature. Actually, $\varepsilon_{\rm AT}$ can be viewed as the minimal fraction of errors for which the replica-symmetric ansatz is physical (stable). This upper bound is presented in figure 2 (upper solid curve). Due to the re-entrance of the de Almeida–Thouless line we find $\varepsilon_{\rm AT} \approx 0.04$ at $\alpha = 2$, so this curve does not appear in the inset of figure 2.

We consider now the first stage of Parisi's replica symmetry breaking scheme. In this case the n replicas are divided into n/m groups of m replicas and one sets $q_{ab} = q_1$, $\phi_{ab} = \phi_1$ if a and b belong to the same group and $q_{ab} = q_0$, $\phi_{ab} = \phi_0$ otherwise. Moreover, $\eta_a = \eta$ for all replicas. This ansatz was studied by Erichsen and Theumann (1993) and the minimal fraction of errors $\varepsilon^{(1)}_{\rm RSB}$ obtained from their equations is shown in figure 2 (short broken curve). As the stability analysis of this solution has not been performed, its validity remains unknown. In this sense, the two bounds presented above may be of utility, since any physical solution should not violate any of them. In fact, the one-step solution satisfies the two bounds for all values of α.

It is interesting to note that, for α not too near 2, the upper bound seems to be closer to the exact solution (one believes that further steps of replica symmetry breaking can only increase the estimate of ε) than the lower bound of Venkatesh and Psaltis.

In summary, we have compared the estimates for the minimal fraction of errors of the replica symmetric and the one-step replica symmetry breaking ansatz with two bounds: a rigorous lower bound demonstrated by Venkatesh and Psaltis (1992) and an exact upper bound obtained in this note. As a consequence, we have verified the tightness of the former bound and provided evidence about the physicality of the replica-symmetry breaking solution.

The authors thank Fudação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS) for supporting the visit of JFF to Porto Alegre. This research was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico ($\rm CNP_q$).

References

de Almeida J R and Thouless D J 1978 *J. Phys. A: Math. Gen.* **11** 983
Brunel N, Nadal J P and Toulouse G 1992 *J. Phys. A: Math. Gen.* **25** 5017
Cover T M 1965 *IEEE Trans. Electron. Comput.* **EC-14** 326
Erichsen Jr R and Theumann W K 1993 *J. Phys. A: Math. Gen.* **26** L61
Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
Parisi G 1980 *J. Phys. A: Math. Gen.* **13** 1101
Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792
Venkatesh S S and Psaltis D 1992 *IEEE Trans. Pattern Analysis and Machine Intelligence* **14** 87