J. Phys. A: Math. Gen. **23** (1990) L935-L938. Printed in the UK

LETTER TO THE EDITOR

On the capacity per synapse

Ido Kanter and Eli Eisenstein

Department of Physics, Bar-Ilan University, Ramat Gan 52100, Israel

Received 18 May 1990

Abstract. The optimal storage capacity of a perceptron with a finite fraction of sign constrained weights, which are prescribed *a priori*, is examined. The storage capacity is calculated by considering the fractional volume of weights which can store a set of αN random patterns, where N is the size of the input. It is found that in the case where $(1-s)N$ weights are sign constrained the capacity is $(1+s)/2\alpha_c^G(k)$, where $\alpha_c^G(k)$ is the maximal storage capacity in Gardner's case and k is a stability parameter.

Numerous analyses of neural networks using statistical mechanics tools were published in the last few years. Analogies between such systems and random magnetic spin systems have been studied extensively [1]. Recently, a new class of networks, so-called multilayer networks, has attracted much attention. The multilayer networks are composed of neurons like binary units interacted in a feedforward fashion, i.e. no loops are allowed in the connected graph. The output of the network can be easily computed by propagating the signals from the input units to the output units. In this process each element is updated according to some transfer function of its input from the previous layer.

The prototype of this class of architectures is the perceptron [2], which consists of N binary input elements and one binary output element. We are interested in the embedding of $p \equiv \alpha N$ relations between pairs of input/output. More precisely, the input of the network is a set of p random patterns $\xi_i^\mu = \pm 1$, $i = 1, \ldots, N$ and $\mu = 1, \ldots, p$ and a set of p random binary outputs $y^\mu = \pm 1$, $\mu = 1, \ldots, p$. For such a network the learning process is to modify the synaptic weights, $\{J_j\}$, in such a way that

$$y^\mu \sum_{j=1}^{N} \frac{J_j}{\sqrt{N}} \xi_j^\mu > \kappa \qquad \mu = 1, \ldots, p. \tag{1}$$

The stability parameter κ is to ensure robustness to errors in the input or to enlarge the basin of attractions in the case of fully connected networks and was first introduced by Gardner [3]. In the case $\kappa > 0$ its value is meaningful only when one specifies the normalization of the weights $\{J_j\}$ [3, 4]. One commonly-used normalization is the spherical normalization

$$\sum_{j=1}^{N} J_j^2 = N \tag{2}$$

which is a global constraint.

A significant contribution to the perceptron problem was carried out by Gardner who showed that the probability existence of solutions can be deduced from the fractional volume of the parameters $\{J_j\}$ which obeys constraints (1) and (2) [3]. Recent

0305-4470/90/170935+04$03.50

work applied this method to various cases and here we would like to emphasize two lines of generalizations. In the first approach, the global constraint is replaced by local constraints on each individual weight. This class of problems contains, for instance, the perceptron in the Ising limit [4, 5] or in the limit of discrete synaptic weights [6]. In the second approach, local constraints are added in addition to the spherical constraint. An example that belongs to this class is a network with sign-constrained synapses [7]. In this case the sign of all the weights are prescribed *a priori*. The study of neural networks with local constraints on the weight strengths is motivated from both biological and applications points of view.

In this work we will concentrate on the second approach, but the local constraints could differ from one weight to another. More precisely, our perceptron consists of sN unconstrained weights

$$J_i \in (-\infty, \infty) \qquad i = 1, \ldots, sN \tag{3}$$

and $(1-s)N$ sign-constrained weights

$$J_i \in (0, \infty) \qquad i = sN+1, \ldots, N. \tag{4}$$

The limit $s = 1$ is Gardner's case, where each point on the sphere, (2), is available as a solution. The second limit $s = 0$ is the sign-constrained weights where only the 2^{-N} connected region of the sphere is available as a solution.

Following Gardner, the relative volume in the weights space for the network defined by (3) and (4) is given by

$$V = C \int_{-\infty}^{\infty} \prod_{j=1}^{sN} \mathrm{d}J_j \int_0^{\infty} \prod_{j>sN} \mathrm{d}J_j \prod_{\mu} \Theta\left(y^{\mu} \sum_j \frac{J_j}{\sqrt{N}} \xi_j^{\mu} - \kappa \right) \delta\left(\sum_j (J_j)^2 - N \right) \tag{5}$$

where C is a normalization constant and the theta and the delta functions stand for constraints (1) and (2) respectively.

The computation proceeds as in [3] and [4]. One concentrates on the computation of $\ln V$ which is a quantity of order N. This quantity is averaged over the quenched distributions of the random input patterns $\{\xi_i^{\mu}\}$, in the expectation that the fluctuations of $\ln V$ from sample to sample are negligible. Using the replica method one should calculate the average over the following quantity

$$\langle V^n \rangle = \left\langle \prod_{\alpha=1}^{n} \int_{-\infty}^{\infty} \prod_{j=1}^{sN} \mathrm{d}J_j^{\alpha} \int_0^{\infty} \prod_{j>sN} \mathrm{d}J_j^{\alpha} \prod_{\mu} \Theta\left(y^{\mu} \sum_j \frac{J_j^{\alpha}}{\sqrt{N}} \xi_j^{\mu} - \kappa \right) \delta\left(\sum_j (J_j^{\alpha})^2 - N \right) \right\rangle \tag{6}$$

where $\langle \ldots \rangle$ stands for the average over the random inputs $\{\xi_i^{\mu}\}$. In the thermodynamic limit and within the replica-symmetric ansatz, (6) can be expressed in terms of three order parameters E, ϕ and q in the following form

$$\langle \ln V \rangle = \mathrm{ext}_{\{q,\phi,E\}} \left[\alpha G_1(q, \kappa) + G_2(\phi, E) + \tfrac{1}{2}\phi q + \tfrac{1}{2}E \right] \tag{7}$$

where

$$G_1(q, \kappa) = \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \ln\left(\int_{(\kappa+\sqrt{q}z)/(\sqrt{1-q})}^{\infty} \frac{\mathrm{d}x}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \right) \tag{8}$$

$$G_2(\phi, E) = \frac{-1}{2} \ln(E+\phi) + \frac{\phi}{2(E+\phi)} + (1-s) \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2} \ln\left(\int_{z\sqrt{\phi/(E+\phi)}}^{\infty} \mathrm{d}J \, \mathrm{e}^{-J^2/2} \right) \tag{9}$$

and the order parameter q is the overlap between two possible solutions represented by weights of two different replicas

$$q \equiv q^{\alpha\beta} = \frac{1}{N}\sum_{j=1}^{N} J_j^{\alpha} J_j^{\beta}. \tag{10}$$

The saddle point equations are obtained by taking the derivatives of $\langle \ln V \rangle$ with respect to q, E and ϕ to zero. Near the critical capacity we assume that $q \to 1$ and $\phi/(E+\phi) \to \infty$. These assumptions will be checked self-consistently to satisfy the solution. In this limit the saddle-point equations with respect to E and ϕ are given respectively by

$$0 = 2(E+\phi)^2 - 2\phi - (1-s)E \tag{11}$$

and

$$0 = q + \frac{(s-1)}{2\phi} + \frac{(s-1)\phi + 2qE}{2(E+\phi)^2}. \tag{12}$$

Near the maximal storage capacity, E and ϕ can be expressed as a power series of $(1-q)$. Substitution of this expansion in equations (11), (12) gives:

$$\phi = \frac{(s-1)}{2(1-q)} + \frac{(s+1)}{2(1-q)^2} \tag{13}$$

and

$$E = \frac{1}{(1-q)} - \frac{(s+1)}{2(1-q)^2}. \tag{14}$$

Now one can indeed verify that as $q \to 1$, $\phi/(E+\phi) \to \infty$. Using these results one can find that (7) can be expressed in the following form

$$\langle \ln V \rangle = \alpha \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \ln\left(\int_{(\kappa+\sqrt{q}z)/(\sqrt{1-q})}^{\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \right) + \frac{(s+1)}{4(1-q)} + \lambda(q) \tag{15}$$

where $\lambda(q)$ is the less singular part of $\langle \ln V \rangle$. In Gardner's case, $s = 1$, this quantity, $\langle \ln V \rangle$, is given by

$$\alpha \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \ln\left(\int_{(\kappa+\sqrt{q}z)/\sqrt{1-q}}^{\infty} \frac{dx}{\sqrt{2\pi}} e^{-x^2/2} \right) + \frac{1}{2(1-q)}. \tag{16}$$

It is now clear that these two equations, (15) and (16), are the same up to rescaling of α. Hence, one can conclude without doing any actual calculations that

$$\alpha_c(s, \kappa) = \frac{(1+s)}{2} \alpha_c(1, \kappa) \tag{17}$$

where $\alpha_c(1, \kappa)$ is the solution in the Gardner's limit, $s = 1$ [3, 4]. In the simple case $\kappa = 0$, $\alpha_c(1, 0) = 2$ and (17) gives

$$\alpha_c(s) = 1 + s \qquad 0 \leqslant s \leqslant 1. \tag{18}$$

Both end points of (18) were previously obtained. The limit $s = 1$ is Gardner's case which gives $\alpha_c = 2$, and the limit $s = 0$ is the sign-constrained case which gives $\alpha_c = 1$ [8].

The results, (17), (18), indicate a general rule. The maximal embedding information in a sign-constrained weight is one half of the maximal embedding information in an unconstrained weight. This result is correct even in the case of positive κ and in the presence of different local constraints on each individual weight. In the simple case, $\kappa = 0$, the maximal embedding information in a constrained weight is one bit of information and the maximal embedding information in the case of unconstrained weight is two bits of information.

It is important to stress that the double maximal embedding information in an unconstrained weight in comparison to a constrained weight where $\kappa = 0$, is a general result which is independent of the possible continuous range for each weight. In order to verify this statement let us examine now a few simple cases.

(*a*) *Ns* weights are constrained to $(-a, a)$ and the rest are constrained to $(0, a)$, where a is a constant. It is obvious that the maximal storage capacity is unchanged, since the weights could be normalized in such a way that $-a \leqslant J \leqslant a$. However, it is also obvious that after the normalization the scale of the spherical constraint is changed.

(*b*) *Ns* weights are constrained to $(-a, a)$ and the rest are constrained to $(0, b)$, where a and b are some constants and we assume that $b > a$. It is obvious that the maximal capacity is bounded from below by the case where *Ns* weights are constrained to $(-a, a)$ and the rest are constrained to $(0, a)$. Using case (*a*), the capacity is bounded from below by $1 + s$. On the other hand, the maximal capacity is bounded from above by the case where *Ns* weights are constrained by $(-b, b)$ and the rest are constrained by $(0, b)$, which gives also $1 + s$. Hence, the maximal storage capacity is again $1 + s$.

(*c*) Using similar ideas, one can verify that for the case $J_i \varepsilon (-a_i, b_i)$, where a_i ($i \leqslant Ns$) and b_i are any positive numbers and a_i is zero for $i > Ns$, the maximal storage capacity is again $1 + s$. The only necessary condition is that the continuous range for each weight includes the origin. Furthermore, the result (18) is unchanged even when one gives a different probability for each possible point in the weights space. In this case, the integrals over the weights, $\int \mathrm{d}J$, in (5) and (6) are replaced by $\int \mathrm{d}J_j \, p_j(J_j)$ where $p_j(J_j)$ is some even positive distribution function. One can carry out the calculation in a similar way and find explicitly for particular cases that the result (18) is unchanged. The idea is that all solutions exist independently of the form of $p(J)$, but with different probabilities. Hence, the maximal storage capacity is not affected by $p(J)$.

Finally, we would like to comment that a similar learning algorithm suggested in reference (8) for Dale's rule can be used for the discussed networks. The stability analysis of this model can also be done but it is expected that the replica symmetric solution is stable, since the available weight space is a connected region.

We thank E Barkai, E Domany, H Gutfreund and J Kurchan for stimulating discussions. The research is supported by the Israeli Ministry of Science and Development.

References

[1] Amit D J 1989 *Modeling Brain Function* (NY: Cambridge University Press)
[2] Minsky M L and Papert S 1969 *Perceptron* (Cambridge, MA: MIT Press)
[3] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[4] Gardner E and Derrida D 1988 *J. Phys. A: Math. Gen.* **21** 271
[5] Krauth W and Mezard M 1989 *J. Physique* **50** 3057
[6] Gutfreund H and Stein Y *J. Phys. A: Math. Gen.* **23** 2613
[7] Amit D J, Wong K Y M and Campbell C 1989 *J. Phys. A: Math. Gen.* **22** 2039
[8] Amit D J, Wong K Y M and Campbell C 1989 *J. Phys. A: Math. Gen.* **22** 4687