Enhancing the learning of a finite number of patterns in neural networks

1988 J. Phys. A: Math. Gen. 21 L253

(http://iopscience.iop.org/0305-4470/21/4/012)

J. Phys. A: Math. Gen. **21** (1988) L253–L257. Printed in the UK

LETTER TO THE EDITOR

Enhancing the learning of a finite number of patterns in neural networks

J F Fontanari and R Köberle

Departamento de Física e Ciência dos Materials, Instituto de Física e Química de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13.560 São Carlos SP, Brazil

Received 6 October 1987

Abstract. We propose a mechanism to enhance the learning of an arbitrary but finite number of patterns (marked patterns) without damaging the ability of the network to retrieve the rest of the patterns. We find that the process of forgetting the marked patterns may become continuous, presenting then an unusual behaviour for symmetric neural networks.

Neural networks with symmetric quadratic interactions have recently become fashionable as models for distributed content-addressable memories. If the states of the neuron are represented by a spin variable S_i, which may assume the values $S_i = +1$ (active) or $S_i = -1$ (passive) the neural network may be viewed as a statistical mechanical system governed by the Hamiltonian (Hopfield 1982):

$$H = -\tfrac{1}{2} \sum J_{ij} S_i S_j. \tag{1}$$

Storage of p input patterns $\{\xi_i^\mu = \pm 1\}$, $\mu = 1, \ldots, p$, is achieved implementing Hebb's rule as

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu. \tag{2}$$

It is assumed that the ξ_i^μ are random independent quenched variables and that the number of memorised patterns (p) is proportional to N:

$$p = \alpha N. \tag{3}$$

The model described above is able to retrieve or recognise the memorised patterns for $\alpha \leqslant \alpha_c \approx 0.14$ with an error smaller than 1.5% (Amit *et al* 1985, 1987, Kinzel 1985).

In this letter we address the problem of enhancing the learning of a finite subset of memories $\{\xi_i^\mu, \mu = 1, \ldots, r\}$. Our motivation to approach this problem comes from biology: biosystems cannot forget information essential for survival. Therefore we must modify the model in such a way that even if the network is overloaded $(\alpha > \alpha_c)$ it will recognise some special patterns.

Enhancing the learning of r memorised patterns (marked patterns) can be achieved by coupling them to an external field (Amit *et al* 1987). However, this approach presents some troubles: the improvement in the retrieval of a marked pattern may preclude the retrieval of all the rest of the memorised patterns. Moreover the efficiency of this mechanism for enhanced learning decreases as the number of marked patterns increases: for $r > 5$ this mechanism becomes useless.

0305-4470/88/040253+05$02.50

To enhance the learning of r patterns we rewrite the learning rule (2) as

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{r}\xi_i^\mu\xi_j^\mu+\frac{\gamma}{N}\sum_{\mu>r}^{p}\xi_i^\mu\xi_j^\mu \tag{4}$$

with $\gamma\leqslant 1$.

Similar modifications of Hebb's rule have been considered recently (Mezard *et al* 1986, Nadal *et al* 1986) in order to avoid overloading and allowing the system to store recent information by forgetting old memories†. Our motivation is different, since we want to retrieve a certain set of memories without degrading the rest.

In the following we will compute the retrieval qualities of the marked and unmarked patterns. We use the replica method to evaluate the quenched free energy

$$-\beta f=\lim_{n\to\infty}\lim_{N\to\infty}\frac{1}{Nn}[\langle Z^n\rangle-1] \tag{5}$$

where

$$Z=\mathrm{Tr}_{S_i}\exp\left[\frac{\beta}{2N}\sum_{\mu=1}^{r}\left(\sum_i\xi_i^\mu S_i\right)^2+\frac{\beta\gamma}{2N}\sum_{\mu>r}^{p}\left(\sum_i\xi_i^\mu S_i\right)^2\right]. \tag{6}$$

Employing standard manipulations (Amit *et al* 1985) we express the quenched free energy in terms of the following order parameters:

(a) the macroscopic overlaps with r marked patterns

$$m^\mu=\frac{1}{N}\left\langle\sum_i\xi_i^\nu\langle S_i\rangle_{\mathrm{T}}\right\rangle_\xi \qquad \mu=1,\ldots,r \tag{7}$$

(b) the macroscopic overlaps with $s-r$ unmarked patterns, s being finite as $N\to\infty$,

$$n^\mu=\frac{1}{N}\left\langle\sum_i\xi_i^\nu\langle S_i\rangle_{\mathrm{T}}\right\rangle_\xi \qquad \mu=r+1,\ldots,s \tag{8}$$

(c) the total mean square of the random overlaps with p-s patterns

$$\tilde{q}=\frac{1}{\alpha N^2}\left\langle\sum_{ij\mu}\xi_i^\mu\xi_j^\mu\langle S_i\rangle_{\mathrm{T}}\langle S_j\rangle_{\mathrm{T}}\right\rangle_\xi \qquad \mu>s \tag{9}$$

(d) the Edwards-Anderson order parameter

$$q=\frac{1}{N}\left\langle\sum_i\langle S_i\rangle_{\mathrm{T}}^2\right\rangle_\xi. \tag{10}$$

Here $\langle\ \rangle_{\mathrm{T}}$ stands for the thermal average and $\langle\ \rangle_\xi$ for the average of the quenched variables $\{\xi_i^\mu\}$.

The free energy per neuron is then given by

$$f=\tfrac{1}{2}\boldsymbol{m}^2+\tfrac{1}{2}\gamma\boldsymbol{n}^2+\tfrac{1}{2}\alpha\beta\gamma^2\tilde{q}(1-q)-\beta^{-1}\langle\!\langle\log(2\cosh(\beta\Xi))\rangle\!\rangle \tag{11}$$

where

$$\Xi=(\alpha\gamma^2\tilde{q})^{1/2}z+\boldsymbol{m}\cdot\boldsymbol{\xi}+\gamma\boldsymbol{n}\cdot\boldsymbol{\xi} \tag{12}$$

† Mezard *et al* (1986) consider the learning rule $J_{ij}=(1/N)\Sigma_\mu^P\Lambda(\mu/N)\xi_i^\mu\xi_j^\mu$ where the function $\Lambda(\mu)$ is suitably normalised. Since our rule is stationary, we relax the normalisation condition as $P\to\infty$, so as to obtain $\Lambda(\mu)=\text{constant}>0$.

and ⟪ ⟫ indicates the average over ξ^μ ($\mu < s$) and over a Gaussian variable z with zero mean and unit variance.

The values of the order parameters are given by the saddle-point equations:

$$m^\mu = \langle\langle \xi^\mu \tanh(\beta\Xi)\rangle\rangle \qquad \mu = 1, \ldots, r \tag{13a}$$

$$n^\mu = \langle\langle \xi^\mu \tanh(\beta\Xi)\rangle\rangle \qquad \mu = r+1, \ldots, s \tag{13b}$$

$$q = \langle\langle \tanh^2(\beta\Xi)\rangle\rangle \tag{13c}$$

$$\tilde{q} = q/[1-\beta\gamma(1-q)]^2. \tag{13d}$$

We shall only study the zero temperature limit of these equations. Firstly we shall examine solutions with a macroscopic overlap with the single marked pattern $\{\xi_i^1\}$:

$$m^\mu = m\delta_{\mu,1} \tag{14a}$$

$$n^\mu = 0 \qquad \mu = r+1, \ldots, s. \tag{14b}$$

As $\beta \to \infty$ the four equations ($13a$–d) can be reduced to one equation for the variable $y = m(2\alpha\gamma^2\tilde{q})^{-1/2}$

$$\gamma y = \operatorname{erf}(y)[(2\alpha)^{1/2} + 2\pi^{1/2}e^{-y^2}]^{-1}. \tag{15}$$

For $y \ll 1$ this equation may be expanded yielding

$$y^2 \approx 3\left(\frac{1+(\alpha\pi/2)^{1/2}}{2-(\alpha\pi/2)^{1/2}}\right)\{\gamma[1+(\alpha\pi/2)^{1/2}]-1\}. \tag{16}$$

Therefore we obtain a continuous transition from a retrieval to a spin-glass phase for

$$\alpha = (2/\pi)(\gamma^{-1}-1)^2 \qquad \gamma^{-1} > 3. \tag{17}$$

This goes over to a discontinuous transition at the tricritical point

$$\alpha_T = 8/\pi \qquad \gamma_T^{-1} = 3 \tag{18}$$

as shown in figure 1(a).

Although the retrieval quality of the marked patterns at $\alpha = \alpha_c(\gamma)$ decreases and goes to zero as γ^{-1} increases (figure 1(b)) it is greatly improved when we keep α fixed and increase γ^{-1} as shown in figure 2.

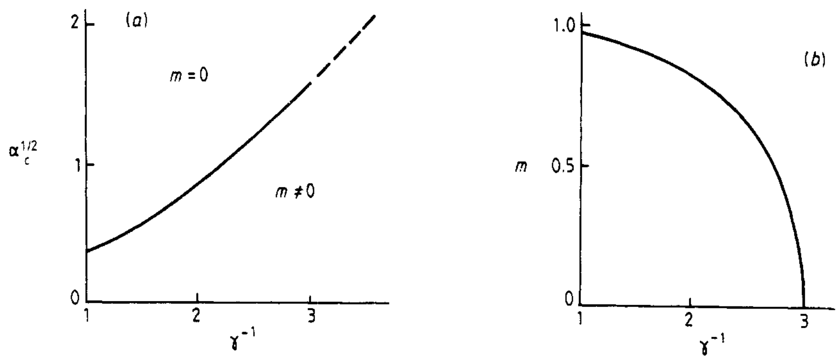

Figure 1. (a) Curve of $\alpha_c^{1/2}(\gamma^{-1})$ below which the marked retrieval states appear. The broken curve represents a continuous transition while the full curve corresponds to a discontinuous transition. (b) The retrieval quality of the marked retrieval states along the curve of (a).

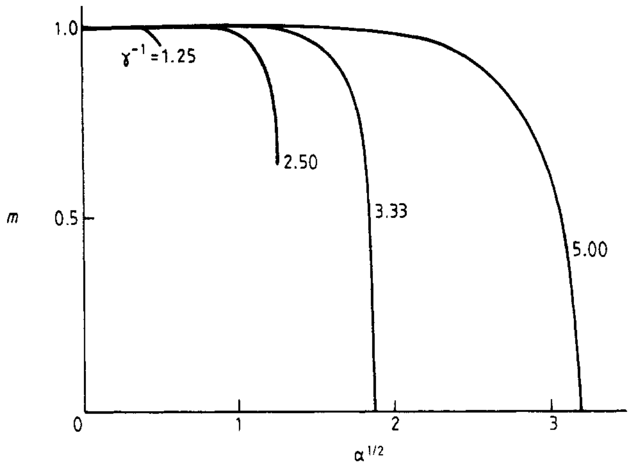

Figure 2. The retrieval quality of the marked retrieval states against $\alpha^{1/2}$ for $\gamma^{-1}=1.25$, 2.50, 3.33 and 5.00. Notice the improvement of m for α fixed as γ^{-1} increases.

It should be emphasised that the increase of $\alpha_c(\gamma)$ does not mean any improvement in the network's storage capacity. It means that it is still possible to retrieve any one of the r marked patterns when there are $p=\alpha_c(\gamma)N$ unmarked patterns stored in the network. Of course, these patterns cannot be retrieved since $\alpha_c(\gamma)>\alpha_c(1)$.

For $\gamma^{-1}>3$ the process of forgetting the marked patterns is continuous. The main advantage of smoothly degrading memories is the possibility of implementing some control mechanism to detect this process. The system then has a chance to avoid the dangerous region where even the marked patterns are not retrievable.

We shall now look at the solutions having macroscopic overlap with the single unmarked pattern $\{\xi_i^\mu\}$:

$$m^\mu=0 \qquad \mu=1,\ldots,r \tag{19a}$$

$$n^\mu=n\delta_{\mu,\mu}. \tag{19b}$$

The equation analogous to (15) is

$$x=\operatorname{erf}(x)[(2\alpha)^{1/2}+2\pi^{-1/2}\,e^{-x^2}]^{-1} \tag{20}$$

where $x=n(2\alpha\tilde{q})^{-1/2}$.

Therefore we get a discontinuous transition from a retrieval to a spin-glass phase for $\alpha_c\approx 0.138$ with $n(\alpha_c)\approx 0.97$ independently of the value of γ.

In conclusion, this letter presents a mechanism for enhancing the learning of an arbitrary but finite number of patterns without damaging the ability of the network to retrieve the rest of the patterns.

The research of RK is partially supported by CNPq and JFF holds a FAPESP fellowship.

Note added in proof. The referee has brought to our attention the work of S Shinomoto (1987 *Biol. Cybern.* **57** 197) where similar issues are addressed.

References

Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1930; 1987 *Ann. Phys., NY* **173** 30
Hopfield J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
Kinzel W Z 1985 *Z. Phys.* B **60** 205
Mezard M, Nadal, J P and Toulouse G 1986 *J. Physique* **47** 1457
Nadal J P, Toulouse G, Changeux J P and S Dehaene 1986 *Europhys. Lett.* **1** 535