Storage capacity of the truncated projection rule

1991 J. Phys. A: Math. Gen. 24 3707

(http://iopscience.iop.org/0305-4470/24/15/036)

J. Phys. A: Math. Gen. 24 (1991) 3707–3709. Printed in the UK

Storage capacity of the truncated projection rule

A Engel† and M Weigt

Institut für Theoretische Physik der Humboldt-Universität, Invalidenstrasse 42, Berlin, 0-1040, Federal Republic of Germany

Received 7 December 1990, in final form 5 April 1991

Abstract. A neural network model storing correlated patterns by using a *local* variant of the projection rule is shown to be equivalent to the Hopfield model.

The projection (pseudo-inverse) learning rule is known to store up to $p = N$ linearly independent patterns $\{\xi_i^\mu\}$; $\xi_i^\mu = \pm 1$; $i = 1, \ldots, N$; $\mu = 1, \ldots, p$ in a network of $N \to \infty$ formal neurons $S_i = \pm 1$ (Kohonen 1984, Personnaz *et al* 1985). According to this rule the synaptic matrix J_{ij} is calculated via

$$J_{ij} = \frac{1}{N} \sum_{\mu,\nu} \xi_i^\mu (\bar{C}^{-1})_{\mu,\nu} \xi_j^\nu \tag{1}$$

where

$$\bar{C}_{\mu\nu} = \frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu \tag{2}$$

is the overlap matrix of the patterns. Although there are fast algorithms to invert $\bar{C}_{\mu\nu}$ the non-locality of (1) is a serious disadvantage in modelling, e.g. learning processes.

Usually one studies ensembles of random patterns with probability distributions factorizing in the neuron index i but not in the pattern index μ, i.e.

$$P(\{\xi_i^1, \xi_i^2, \ldots, \xi_i^p\}) = \prod_i P(\xi_i^1, \xi_i^2, \ldots, \xi_i^p). \tag{3}$$

Correlations between the patterns giving rise to non-trivial forms of the overlap matrix $\bar{C}_{\mu\nu}$ occur due to correlations between the values of different patterns at the same neuron. Well-known examples are patterns with low levels of activity and hierarchically correlated patterns.

The matrix

$$C_{\mu\nu} = \langle\!\langle \xi_i^\mu \xi_i^\nu \rangle\!\rangle$$

where $\langle\!\langle \ldots \rangle\!\rangle$ denotes the average over the distribution (3), then differs from $\bar{C}_{\mu\nu}$ by terms of order $N^{-1/2}$ only. If one were using $C_{\mu\nu}^{-1}$ instead of $\bar{C}_{\mu\nu}^{-1}$ in (1) the resulting learning rule would be local, since $C_{\mu\nu}$ depends only on the properties of the ensemble of patterns and not on the particular realization.

† Present address: Institut für Theoretische Physik der Universität Göttingen, Bunsenstrasse 9, 3400 Göttingen, Federal Republic of Germany.

0305-4470/91/153707+03$03.50

In fact this replacement was used recently for the special case of a regular pattern hierarchy (Cortes *et al* 1987). Somewhat surprisingly, the resulting model turned out to be equivalent to the Hopfield model using the Hebb rule and correspondingly the storage capacity was $\alpha_c = 0.14$ instead of $\alpha_c = 1$.

In this paper we show that this result is rather general. For arbitrary pattern statistics with the property (3) the 'truncated' projection rule

$$J_{ij} = \frac{1}{N} \sum_{\mu,\nu} \xi_i^\mu (C^{-1})_{\mu\nu} \xi_j^\nu \tag{4}$$

can store up to $p_{\max} = 0.14\ N$ patterns. Moreover, the thermodynamic properties of the phases with one condensed pattern are exactly the same as in the Hopfield model (Amit *et al* 1987).

To prove this we construct a linear transformation $T_{\mu\nu}$ to a new set of patterns $\{\zeta_i^\mu\}$. The original pattern set $\{\xi_i^\mu\}$ is characterized by

$$\xi_i^\mu = \pm 1 \qquad \langle\langle \xi_i^\mu \rangle\rangle =: c^\mu \qquad \langle\langle \xi_i^\mu \xi_i^\nu \rangle\rangle =: C^{\mu\nu}.$$

For the transformation

$$\zeta_i^\mu = \sum_\nu T_{\mu\nu} \xi_i^\nu$$

we require

$$\zeta_i^1 = \xi_i^1 \tag{5}$$

and

$$\langle\langle \zeta_i^\mu \rangle\rangle = 0 \qquad \langle\langle \zeta_i^\mu \zeta_i^1 \rangle\rangle = 0 \tag{6}$$

for as many $\mu \geqslant 2$ as possible.

Equations (6) give rise to two equations for the rows of $T_{\mu\nu}$. Since we are looking for a *regular* transformation the rows must be linearly independent of each other and hence (6) can be fulfilled for $(p-2)$ patterns $\{\zeta_i^3\}, \ldots, \{\zeta_i^p\}$. With the help of an appropriate orthogonalization procedure we can therefore find a transformation which gives rise to

$$\begin{aligned} &\langle\langle \zeta^1 \rangle\rangle = c^1 =: c && \langle\langle \zeta^2 \rangle\rangle =: b \neq 0 \\ &\langle\langle \zeta^\mu \rangle\rangle = 0 && \mu \geqslant 3 \\ &\langle\langle \zeta^\mu \zeta^\nu \rangle\rangle = \delta^{\mu,\nu} && \mu, \nu = 1, \ldots, p. \end{aligned} \tag{7}$$

From (4) and (7) we then find for the synaptic couplings

$$J_{ij} = \frac{1}{N} \sum_\mu \zeta_i^\mu \zeta_j^\mu. \tag{8}$$

The free energy is now calculated using standard techniques [4]. Note that only the condensed pattern $\{\zeta_i^1\}$ has the usual binary distribution

$$P(\zeta_i^1) = \frac{1+c}{2} \delta(\zeta_i^1 - 1) + \frac{1-c}{2} \delta(\zeta_i^1 + 1)$$

because of (5). For $\mu \geqslant 2$ the ζ_i^μ are not restricted to the values ± 1. However, due to the central limit theorem for the average over these 'high' patterns only the first two moments are needed (see Amit *et al* 1987) which are given by (7). Hence the average over the ζ_i^μ with $\mu \geq 3$ gives the same result as for the Hopfield model. For ζ_i^2 we use the distribution

$$P(\zeta_i^2) = (2\pi)^{-1/2} \exp\left[-\frac{1}{2}\left(\zeta_i^2 - \frac{b}{1-c^2} + \frac{bc}{1-c^2}\zeta_i^1\right)^2\right]$$

in order to meet (7). The calculation of the free energy can be performed within the replica-symmetric approximation by introducing the usual order parameters m and q and in addition $a = (1/N)\sum S_i$ together with their conjugated Lagrange multipliers l, r and k respectively. For the self-consistent equations we get

$$\begin{aligned}
m &= \langle\langle \zeta^1 \tanh \beta(k + l\zeta^1 + (\alpha r)^{1/2} z)\rangle\rangle \\
l &= m - \frac{2b^2 c(a - cm)}{(1-c^2)^2(1-\beta+\beta q)} \\
a &= \langle\langle \tanh \beta(k + l\zeta^1 + (\alpha r)^{1/2} z)\rangle\rangle \\
k &= \frac{2b^2(a - cm)}{(1-c^2)^2(1-\beta+\beta q)} \\
q &= \langle\langle \tanh^2 \beta(k + l\zeta^1 + (\alpha r)^{1/2} z)\rangle\rangle \\
r &= (1-\beta+\beta q)^{-2}\left(q + \frac{2b^2(a-cm)^2}{\alpha(1-c^2)^2}\right).
\end{aligned} \tag{9}$$

As usual $\langle\langle \ldots \rangle\rangle$ now denotes the average over ζ^1 and a Gaussian variable z with zero mean and unit variance. Equations (9) are solved by $a = cm$ giving rise to $l = m$ and $k = 0$. The remaining equations for m, q and r are exactly the same as those found by Amit *et al* for the Hopfield model. Note that the bias c in the ζ^1 average is irrelevant.

It is hence possible for pattern ensembles with site-factorizing but otherwise arbitrary statistics to construct a *local* learning rule with a similar performance as the Hebb rule for independent, non-biased patterns. On the one hand this underlines the universality of the storage capacity $\alpha_c = 0.14$. On the other hand it indicates that the improvement to $\alpha_c = 1$ (Kanter and Sompolinsky 1987) is just due to the $O(N^{-1/2})$ differences between $\bar{C}_{\mu\nu}$ and $C_{\mu\nu}$.

It should be noted that the proposed learning rule is local in so far as the value of the synapse J_{ij} is determined by information about the pattern set at neurons i and j only. Nevertheless, adding a new pattern to the set of stored patterns is more cumbersome than for the Hebb rule since one needs the values of *all* patterns at neurons i and j in order to determine the new value of J_{ij}, whereas for the Hebb rule just the values of the *new* pattern at these neurons suffices. Still, the proposed learning rule is superior to the original projection rule in the sense that the value of J_{ij} remains unaffected by changes of some patterns at other neurons k.

References

Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
Cortes C, Krogh A and Hertz J A 1987 *J. Phys. A: Math. Gen.* **20** 4449
Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
Kohonen T 1984 *Self-Organization and Associative Memory* (Berlin: Springer)
Personnaz L, Guyon I and Dreyfus G 1985 *J. Phys. Lett.* **46** 359