Categorization in the pseudo-inverse neural network

1998 J. Phys. A: Math. Gen. 31 531

(http://iopscience.iop.org/0305-4470/31/2/013)

J. Phys. A: Math. Gen. **31** (1998) 531–540. Printed in the UK PII: S0305-4470(98)84868-3

Categorization in the pseudo-inverse neural network

C Rodrigues Neto and J F Fontanari

Instituto de Física de São Carlos, Universidade de São Paulo, Caixa Postal 369, 13560-970 São Carlos SP, Brazil

Received 9 June 1997, in final form 8 October 1997

Abstract. We investigate analytically the emergence of the categorization ability in the pseudo-inverse attractor neural network. More pointedly, we consider the problem of learning an extensive number of concepts αN by storing a finite number of examples s of each concept. We find that there is a critical value $s_c = 1/\alpha$ beyond which the categorization error, as measured by the average fraction of unstable sites in the concepts, decreases monotonically with s.

1. Introduction

The research on attractor neural networks has focused mainly on their retrieval properties due to the prospective use of these systems as associative memory devices. The basic idea due to Hopfield [1] is to adjust the synaptic couplings J_{ij} between the N neurons that compose the network so that a prespecified set of binary patterns become the attractors of the neural dynamics

$$S_i(t+1) = \text{sign}\left(\sum_j J_{ij} S_j(t) \right) \qquad i = 1, \ldots, N. \tag{1}$$

Here the spin $S_i(t) = \pm 1$ describes the state of the neuron i at time t. Another interesting emergent property of attractor neural networks is the categorization capability, namely, the ability to create a representation (i.e. an attractor) for a concept to which the network has been exposed only through examples [2]. Although this property has been fully investigated for the Hopfield model [2–5], it is not known whether other models of associative memory exhibit this ability as well.

The aim of this paper is to study the categorization ability of the pseudo-inverse attractor neural network [6, 7]. Within the framework proposed by one of us [2], we consider the problem of storing sP binary patterns $\boldsymbol{\xi}^{kl} = (\xi_1^{kl}, \ldots, \xi_N^{kl})$, with $k = 1, \ldots, P$ and $l = 1, \ldots, s$, whose components are generated by the conditional probability distribution

$$p(\xi_i^{kl} | \xi_i^k) = \frac{(1+b)}{2} \delta(\xi_i^{kl} - \xi_i^k) + \frac{(1-b)}{2} \delta(\xi_i^{kl} + \xi_i^k) \tag{2}$$

where the pattern $\boldsymbol{\xi}^k = (\xi_1^k, \ldots, \xi_N^k)$ is the concept underlying the s examples $\boldsymbol{\xi}^{kl}$. Here the parameter $b \in [0, 1]$ measures the average overlap between the examples and their corresponding concept. It is related to the average normalized Hamming distance d between these patterns by $d = (1-b)/2$. As usual, the components of the concepts ξ_i^k are chosen as ± 1 with equal probability. Moreover, we assume that the number of concepts scales linearly with N, i.e. $P = \alpha N$, and that the number of examples s of each concept remains finite in the thermodynamic limit.

0305-4470/98/020531+10$19.50

In the regime where the sP examples are stored perfectly ($sP \leqslant N$) there is an analytical prescription for writing the synaptic couplings of the pseudo-inverse in terms of the stored patterns [6, 7],

$$J_{ij} = \frac{1}{N} \sum_{kl} \sum_{k'l'} \xi_i^{kl} \xi_j^{k'l'} (\mathcal{C}^{-1})_{kl,k'l'} \tag{3}$$

where $\mathcal{C}$ is the correlation matrix whose elements are $\mathcal{C}_{kl,k'l'} = 1/N \sum_i \xi_i^{kl} \xi_i^{k'l'}$. The storage capability and the retrieval properties of a variant of this model, where the diagonal terms are set to zero, was studied in the case where $b = s = 1$ [8]. In this paper, however, we use an alternative formulation which allows us to also investigate the regime of imperfect storage. This is achieved by calculating the minimal norm solution of the following set of sP linear equations [9]

$$\Delta_i^{kl} = 1 \qquad k = 1, \ldots, P \qquad \text{and} \qquad l = 1, \ldots, s \tag{4}$$

for each $i = 1, \ldots, N$, where

$$\Delta_i^{kl} = \frac{1}{\sqrt{N}} \xi_i^{kl} \sum_{j \neq i} J_{ij} \xi_j^{kl} \tag{5}$$

is termed the stability of the example component ξ_i^{kl}. In contrast to equation (3), this prescription yields an asymmetric coupling matrix. Since it does not specify the diagonal terms, we impose the additional condition $J_{ii} = 0$. Hence its storage capacity is $\alpha_c = 1/s - 1/sN \approx 1/s$. In the regime of imperfect storage ($\alpha > \alpha_c$) the pseudo-inverse solution is given by the set of couplings that minimize the following quadratic form, called training energy,

$$E_i(J_{ij}) = \tfrac{1}{2} \sum_{kl} (1 - \Delta_i^{kl})^2 \tag{6}$$

for $i = 1, \ldots, N$. This formulation of the pseudo-inverse has been used to investigate the parallel dynamics of extremely diluted [9] as well as fully connected [10] attractor neural networks in the case where $b = s = 1$. We must emphasize that the previous analyses of the pseudo-inverse attractor neural network [8–10, 18] have concentrated on its retrieval properties, while in this paper we focus on a totally different emergent property, namely, the categorization ability. Moreover, those analyses were restricted to the regime of perfect storage, while the main concern of the present study is the regime of imperfect storage ($\alpha > \alpha_c$), where the categorization property is expected to emerge. In fact, the analytical study of this regime is quite challenging, since neither the thermodynamics approach [8] nor the diluted approximation [9, 11, 18] can be used. The former because there is no analytical prescription for the couplings, and the latter because the examples are correlated, i.e. $\langle \xi_i^{kl} \xi_i^{kl'} \rangle_\xi = b^2$ for $l \neq l'$. Here, $\langle \ldots \rangle_\xi$ stands for the averages over the sets of patterns $\{\boldsymbol{\xi}^{kl}\}$ and $\{\boldsymbol{\xi}^{k}\}$.

The statistical properties of the ensemble of couplings that satisfy (4) or minimize (6) can be investigated analytically using the statistical mechanics formalism developed by Gardner [12, 13]. In this paper we employ that formalism to calculate the average categorization error ϵ_c defined by

$$\epsilon_c = \left\langle\!\!\left\langle \frac{1}{PN} \sum_{k=1}^{P} \sum_{i=1}^{N} \Theta(-\Delta_i^k) \right\rangle_J \right\rangle_\xi \tag{7}$$

where $\Theta(x) = 1$ if $x \geqslant 0$ and 0 otherwise, and

$$\Delta_i^k = \frac{1}{\sqrt{N}} \xi_i^k \sum_{j \neq i} J_{ij} \xi_j^k \tag{8}$$

is the stability of the concept component ξ_i^k. The notation $\langle \ldots \rangle_J$ is the thermal average, i.e. the average over the ensemble of couplings that satisfy equation (4). For sake of completeness we also calculate the average storage or classification error ϵ_s defined by

$$\epsilon_s = \left\langle\!\!\left\langle \frac{1}{sPN} \sum_{k=1}^{P} \sum_{l=1}^{s} \sum_{i=1}^{N} \Theta(-\Delta_i^{kl}) \right\rangle_J \right\rangle_\xi . \tag{9}$$

We note that ϵ_c and ϵ_s give the average fraction of unstable sites in the concepts and examples, respectively. Actually, the statistical independence of the components ξ_i^{kl} and ξ_i^k for different sites, examples, and concepts allows us to evaluate ϵ_c using an alternative formulation. First we calculate the probability distribution of the stabilities of the concepts, namely,

$$\mathcal{W}_c(\gamma) = \langle\!\langle \delta(\gamma - \Delta_i^k) \rangle_J \rangle_\xi \tag{10}$$

and then evaluate ϵ_c using the relation

$$\epsilon_c = \int_{-\infty}^{0} d\gamma \, \mathcal{W}_c(\gamma). \tag{11}$$

Clearly, since the components ξ_i^k are equivalent, $\mathcal{W}_c$ is independent of the indices i and k. A similar procedure can be used to calculate the storage error ϵ_s.

The remainder of the paper is organized as follows. In section 2 we present the calculation of the free-energy and order parameters that characterize the ensemble of couplings of the pseudo-inverse. In section 3 we calculate the average categorization error and in section 4 the average storage error. Section 5 is devoted to our concluding remarks.

2. Free energy

For a fixed site i, the couplings that satisfy (4) are the global minima of the training energy (6). The properties of these minima can be investigated using standard statistical mechanics techniques [12, 13]. More specifically, we focus on the zero-temperature limit of the average free-energy per site

$$f_i = -\frac{1}{N\beta} \langle \ln Z_i \rangle_\xi \tag{12}$$

where Z_i is the partition function

$$Z_i = \prod_j \int_{-\infty}^{\infty} dJ_{ij} \, \delta\left(Q_i - \frac{1}{N} \sum_j J_{ij}^2 \right) \exp[-\beta E_i(J_{ij})]. \tag{13}$$

Here β is the inverse temperature, and $Q_i = Q$ is the squared norm of the couplings which must be carefully chosen in order to give the pseudo-inverse solution [9, 14]. As usual, the evaluation of the quenched average in equation (12) can be effectuated through the replica method: using the identity

$$\langle \ln Z_i \rangle_\xi = \lim_{n \to 0} \frac{1}{n} \ln \langle Z_i^n \rangle_\xi \tag{14}$$

we first calculate $\langle Z_i^n \rangle_\xi$ for *integer* n and then analytically continue to $n = 0$ [15, 16]. Fortunately, since this model is a variant of the spherical model of a spin glass [17], the replica symmetric ansatz for the order parameters, besides being locally stable [14], is probably the exact solution to the problem of extremizing the free-energy density. As the

calculations are straightforward and rather unilluminating we present only the final result for the replica-symmetric free-energy density:

$$-\beta f_i = \frac{1}{2}\left[1+\ln\left(\frac{\pi}{2}\right)\right] + \frac{1}{2}\ln(Q-q_i) + \frac{q_i}{2(Q-q_i)} - \frac{\alpha}{2}(s-1)\ln[1+\beta(Q-q_i)(1-b^2)] - \frac{\alpha}{2}\ln[1+\beta(Q-q_i)(1-b^2+sb^2)] - \frac{\alpha}{2}\frac{\beta s(1+q_i)[1+\beta(Q-q_i)(1-b^2)(1-b^2+sb^2)]}{[1+\beta(Q-q_i)(1-b^2)][1+\beta(Q-q_i)(1-b^2+sb^2)]} \tag{15}$$

where $q_i = \frac{1}{N}\langle\sum_j \langle J_{ij}\rangle_J^2\rangle_\xi$ is the Edwards–Anderson order parameter which must be chosen so as to maximize f_i. Clearly, since the different sites i are equivalent we can set $q_i = q$ and $f_i = f$. In order to single out the states that minimize the training energy (6) we must take the limit $\beta \to \infty$ in equation (15). We note that in this limit $f \to \epsilon$, where ϵ is the average training error. In the regime $q < Q$, the limit $\beta \to \infty$ in the saddle-point equation $\partial f/\partial q = 0$ yields

$$q = \frac{\alpha s}{1-\alpha s} \tag{16}$$

so that $\epsilon = 0$. Since $q \geqslant 0$ this regime of perfect storage exists only for $\alpha \leqslant \alpha_c = 1/s$. In this case the minimal value that the squared norm Q can take on is

$$Q = q = \frac{\alpha s}{1-\alpha s}. \tag{17}$$

To study the regime $\alpha > 1/s$ we must take the limits $\beta \to \infty$ and $q \to Q$ such that $x \equiv \beta(Q-q) > 0$ [13]. This procedure yields the following average training error

$$\epsilon = -\frac{Q}{2x} + \frac{\alpha}{2}\frac{s(1+Q)[1+x(1-b^2)(1-b^2+sb^2)]}{[1+x(1-b^2)][1+x(1-b^2+sb^2)]}. \tag{18}$$

In this case both parameters Q and x are obtained by solving the saddle-point equations $\partial\epsilon/\partial Q = 0$ and $\partial\epsilon/\partial x = 0$. In particular, the former equation yields a quadratic equation for x,

$$x^2(\alpha s-1)(1-b^2+sb^2)(1-b^2) + x[\alpha s - 2(1-b^2) - sb^2] - 1 = 0. \tag{19}$$

Inserting the expression of αs obtained from this equation into the saddle-point equation $\partial\epsilon/\partial x = 0$ yields

$$Q = \frac{xC}{1-xC} \tag{20}$$

where

$$C = \frac{[1+x(1-b^2)(1-b^2+sb^2)]^2 + b^4(s-1)}{[1+x(1-b^2+sb^2)(1-b^2)][1+x(1-b^2)][1+x(1-b^2+sb^2)]}. \tag{21}$$

Thus, given the control parameters α, s and b, we first obtain x by solving equation (19) and then use equation (20) to find Q. We note that we must choose the root of (19) that gives the largest training error ϵ [15, 16]. For $s \to \infty$ we find that $x \to 0$ like $1/s$, while Q remains non-zero.

Since in the context of categorization the important parameter is the number of examples per category s, rather than the number of concepts αN, henceforth we will refer to the transition from the regime of perfect storage ($\epsilon = 0$) to the regime of imperfect storage ($\epsilon > 0$) as occurring at $s = s_c$ where

$$s_c = \frac{1}{\alpha}. \tag{22}$$

3. Categorization error

As mentioned in the introduction, to evaluate the average categorization error we will first calculate the probability distribution of the stabilities Δ_i^k. This is achieved by introducing the auxiliary training energy

$$E_i^c(J_{ij}) = \frac{1}{2}\sum_{kl}(1-\Delta_i^{kl})^2 + \frac{h}{P}\sum_k \delta(\gamma - \Delta_i^k) \tag{23}$$

with Δ_i^{kl} and Δ_i^k given by (5) and (8), respectively. Hence

$$\begin{aligned}\mathcal{W}_c(\gamma) &= \langle\langle \delta(\gamma - \Delta_i^k)\rangle_J\rangle_\xi \\ &= \left\langle\left\langle \frac{1}{P}\sum_k \delta(\gamma - \Delta_i^k)\right\rangle_J\right\rangle_\xi \\ &= -\lim_{\beta\to\infty}\frac{1}{\beta}\frac{\partial\langle \ln Z_i^c\rangle_\xi}{\partial h}\bigg|_{h=0}\end{aligned} \tag{24}$$

where Z_i^c is the partition function (13) with E_i replaced by E_i^c. The calculation of $\mathcal{W}_c$ is very tedious and presents no particular difficulty since it involves the evaluation of Gaussian integrals only. The final result is

$$\mathcal{W}_c(\gamma) = \langle \omega_c(\gamma,\eta)\rangle_\eta \tag{25}$$

where

$$\omega_c(\gamma,\eta) = [2\pi\sigma_c^2]^{-1/2}\exp\left[-\frac{(\gamma - \overline{\gamma_c}(\eta))^2}{2\sigma_c^2}\right] \tag{26}$$

and the notation $\langle\ldots\rangle_\eta$ stands for the average over the random variable $\eta \equiv 2m-s$ distributed according to the binomial probability distribution

$$B(m;s,b) = \frac{s!}{m!(s-m)!}\left(\frac{1+b}{2}\right)^m\left(\frac{1-b}{2}\right)^{s-m} \qquad m = 0,1,\ldots,s. \tag{27}$$

In the regime of perfect storage $s \leqslant s_c$, the mean and the variance of the Gaussian distribution ω_c are given by

$$\overline{\gamma_c}(\eta) = \frac{b\eta}{1-b^2+sb^2} \tag{28}$$

and

$$\sigma_c^2 = \frac{Q(1-b^2)}{1-b^2+sb^2} \tag{29}$$

with Q given by equation (17). We note that for $s = 1$, $\mathcal{W}_c$ gives the probability distribution of the stabilities of a test pattern $\boldsymbol{\xi}^k$ that is at a fixed Hamming distance d to the fixed point $\boldsymbol{\xi}^{k1}$. The analysis of the number of unstable sites of the test pattern was used to investigate the nature of the neighbourhood of the fixed points in the pseudo-inverse as well as in the optimal weights attractor neural networks [18].

In the regime of imperfect storage $s > s_c$ we find

$$\overline{\gamma_c}(\eta) = \frac{xb\eta}{1+x(1-b^2+sb^2)} \tag{30}$$

and

$$\sigma_c^2 = \frac{Q[1+2x(1-b^2)+x^2(1-b^2)(1-b^2+sb^2)]}{[1+x(1-b^2+sb^2)]^2} \tag{31}$$

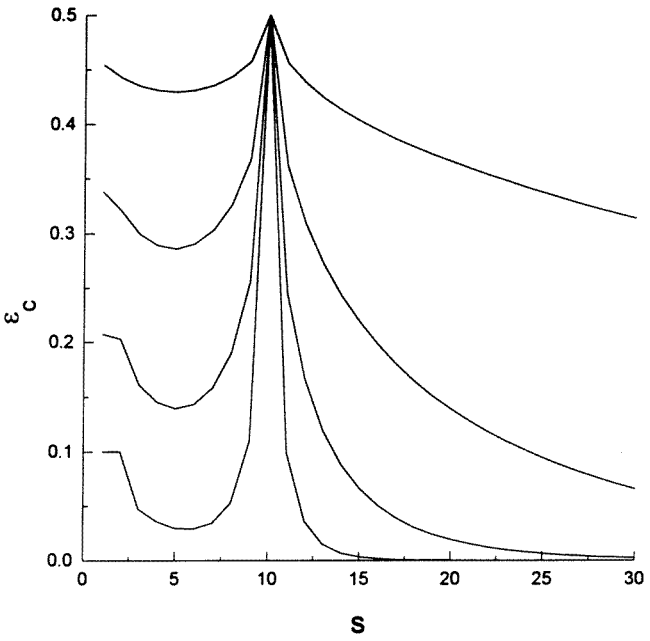

Figure 1. Average categorization error as a function of the number of examples per category for $\alpha = 0.1$ and (from bottom to top) $d = 0.1$, 0.2, 0.3 and 0.4. The examples are stored without error for $s \leqslant 10$.

with Q and x given by equations (20) and (19), respectively.

With these results the evaluation of the categorization error given by equation (11) is straightforward and yields

$$\epsilon_c = \frac{1}{2}\sum_{m=0}^{s} B(m; s, b)\,\mathrm{erfc}\left[\frac{\overline{\gamma_c}(\eta)}{\sqrt{2\sigma_c^2}}\right]. \tag{32}$$

In figure 1 we present the average categorization error as a function of the number of examples per category for fixed α and several values of the average Hamming distance d between the examples and their corresponding concept. The divergence of Q_i, given in (17) and (20), at $s = s_c$ for any $d > 0$ is a typical feature of the pseudo-inverse learning rule [9, 14], which leads to the divergence of σ_c^2 and hence to the result $\epsilon_c = \frac{1}{2}$. This result can be understood by noting that in a network with such large couplings tuned to satisfy (4), the flipping of any extensive number of sites of the stable pattern $\boldsymbol{\xi}^{kl}$, say $j = 1, \ldots, dN$ with $0 < d \leqslant \frac{1}{2}$, will very likely unstabilize all sites i for which the sum $\xi_i^{kl}\sum_{j=1}^{dN} J_{ij}\xi_j^{kl}$ is positive, as this term is almost certainly larger than 1 for any d. Moreover, due to the random character of the components ξ_i^{kl} and the need to satisfy the constraint (4), on average this term is positive for half of the sites only, which then become unstable, leading to $\epsilon_c = \frac{1}{2}$.

For some values of α there is a remarkable similarity between the results presented in figure 1 and those obtained for the extremely diluted Hopfield model [5]. (We note that for $\alpha > 0$ the retrieval regime is absent in the fully connected Hopfield model [2, 4].) Similarly to the Hopfield model at non-zero α, for $s > s_c$ the categorization error decreases monotonically with s and tends to a non-zero asymptotic value ϵ_c^∞ in the limit $s \to \infty$, indicating thus that the concepts never become stable patterns. For $\alpha \leqslant 1$, this result may

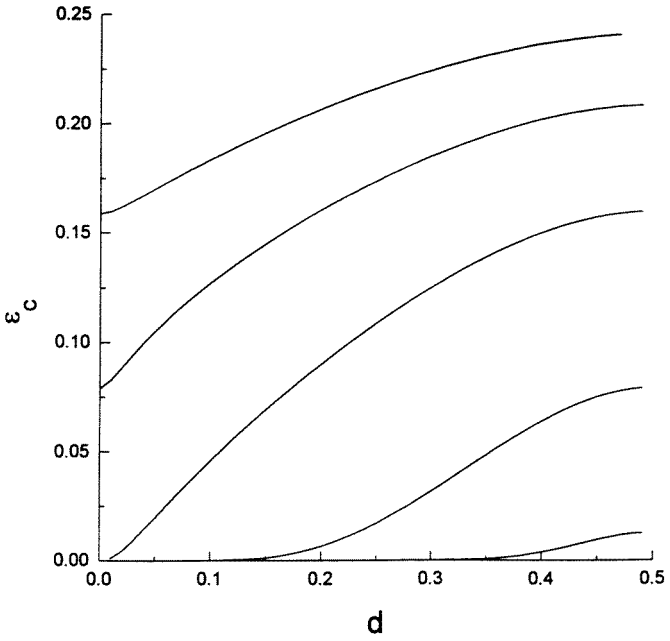

Figure 2. Asymptotic value of the average categorization error as a function of the normalized Hamming distance d between examples and concepts for (from bottom to top) $\alpha = 0.2$, 0.5, 1, 1.5 and 2.

seem quite surprising since one expects naively that for $s \to \infty$ the randomness in the examples will somehow be averaged out, and so the network will effectively store αN patterns, namely, the concepts. Hence, as $\alpha \leqslant 1$, the concepts should be stable patterns. In fact, this is exactly what happens in the Hopfield model [2], except that the stored patterns are never stable in that model. Of course, the problem with this reasoning is that in the Hopfield model the randomness in the examples is averaged out because of the additive character of the Hebb learning rule [1], while the pseudo-inverse rule is set up precisely to avoid the averaging out of the stored patterns, otherwise it would be useless to store correlated patterns [6, 7]. Thus the finding that ϵ_c^∞ is non-zero for the pseudo-inverse is a non-trivial one. In figure 2 we show ϵ_c^∞ as a function of d for several values of α. Note that for $d = 0$ the categorization error coincides with the storage error and hence it vanishes for $\alpha \leqslant 1$. In particular, for $d \approx 0$ we find

$$\epsilon_c^\infty \approx \frac{1}{2}\operatorname{erfc}\left[\frac{1}{\sqrt{2(\alpha-1)}}\right] \qquad \alpha > 1 \tag{33}$$

and

$$\begin{aligned} \epsilon_c^\infty &\approx \tfrac{1}{2}\operatorname{erfc}(\delta) \qquad \alpha \leqslant 1 \\ &\approx \frac{1}{2\sqrt{\pi}\delta}\exp(-\delta^2) \qquad \alpha < 1 \end{aligned} \tag{34}$$

where $\delta = (s_c - 1)^{3/2}/d\sqrt{32}$.

In figure 3 we present the average categorization error as a function of α. It is interesting to note that for fixed s and d there is an optimal value of $\alpha > 1/s$ so that ϵ_c is minimized. Hence a moderate amount of static noise, caused by the microscopic overlap between the

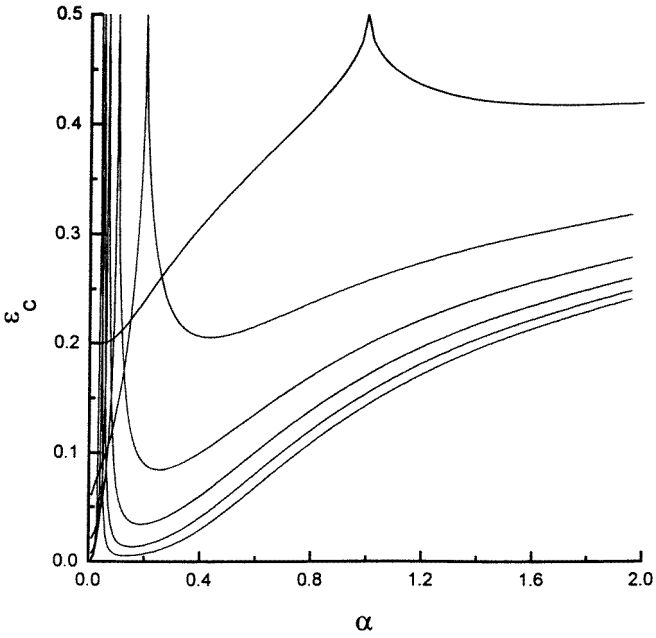

Figure 3. Average categorization error as a function of α for $d = 0.2$ and (from bottom to top at $\alpha = 2$) $s = 25$, 20, 15, 10, 5, and 1. The examples are stored without error for $s \leqslant 1/\alpha$.

concepts, actually improves the categorization ability. Similar results were reported for the Hopfield model in the cases of fast noise (non-zero temperature) [4] and static noise caused by dilution [5]. As already mentioned, the peaks at $\alpha = \alpha_c = 1/s$ are due to the divergence of Q. For $\alpha \to \infty$ all curves tend to $\epsilon_c = \frac{1}{2}$.

4. Storage error

The relevant quantity for the analysis of the storage properties of the pseudo-inverse attractor neural network is not the average training error ϵ, which measures the quadratic deviation of Δ_i^{kl} from 1, but the average classification or storage error ϵ_s since it is only the sign of Δ_i^{kl} that matters for the stability of the component ξ_i^{kl}. Clearly, if $\epsilon = 0$ then $\epsilon_s = 0$. The procedure to calculate the average storage error is the same used in the previous section. More pointedly, we introduce the auxiliary training energy

$$E_i^s(J_{ij}) = \frac{1}{2}\sum_{kl}(1 - \Delta_i^{kl})^2 + \frac{h}{sP}\sum_{kl}\delta(\gamma - \Delta_i^{kl}) \tag{35}$$

with Δ_i^{kl} given by (5), so that

$$\begin{aligned}\mathcal{W}_s(\gamma) &= \langle\langle \delta(\gamma - \Delta_i^{kl})\rangle_J\rangle_\xi \\ &= \left\langle\left\langle \frac{1}{sP}\sum_{kl}\delta(\gamma - \Delta_i^{kl})\right\rangle_J\right\rangle_\xi \\ &= -\lim_{\beta\to\infty}\frac{1}{\beta}\frac{\partial\langle \ln Z_i^s\rangle_\xi}{\partial h}\bigg|_{h=0}.\end{aligned} \tag{36}$$

Here Z_i^s is the partition function (13) with E_i replaced by E_i^s. Although the calculation of $\mathcal{W}_s$ is straightforward, in the sense that it involves only the evaluation of Gaussian integrals,

it proved to be much more involved than the calculation of $\mathcal{W}_c$. In the following we present the final results only.

As before, $\mathcal{W}_s$ can be written as

$$\mathcal{W}_s(\gamma) = \langle \omega_s(\gamma, \eta) \rangle_\eta \tag{37}$$

where $\omega_s(\gamma, \eta)$ is a Gaussian distribution of mean $\overline{\gamma_s}(\eta)$ and variance $\sigma_s^2(\eta)$. In the regime of perfect storage $s \leqslant s_c$ we find $\overline{\gamma_s}(\eta) = 1$ and $\sigma_s^2(\eta) = 0$, so that $\mathcal{W}_s(\gamma) = \delta(\gamma - 1)$. Hence, as expected, $\epsilon_s = 0$ in this regime. For $s > s_c$ we find

$$\overline{\gamma_s}(\eta) = \frac{sx(1-b^2)[1+x(1-b^2+sb^2)] + xb^2\eta^2}{s[1+x(1-b^2)][1+x(1-b^2+sb^2)]} \tag{38}$$

and

$$\sigma_s^2(\eta) = \frac{s^2 A_Q Q + b^2 x^2 A_\eta \eta^2}{s^2[1+x(1-b^2)]^2[1+x(1-b^2+sb^2)]^2} \tag{39}$$

where

$$A_Q = 1 + x(1-b^2+sb^2)(1-b^2)[2+x(1-2b^2+sb^2)] \tag{40}$$

and

$$A_\eta = s(1-b^2+sb^2)[1-x(1-b^2)] - 2s(1-b^2) - b^2\eta^2. \tag{41}$$

Here the order parameters Q and x are given by equations (20) and (19), respectively. Thus the average storage error for $s > s_c$ is given by

$$\epsilon_s = \int_{-\infty}^{0} d\gamma \, \mathcal{W}_s(\gamma)$$

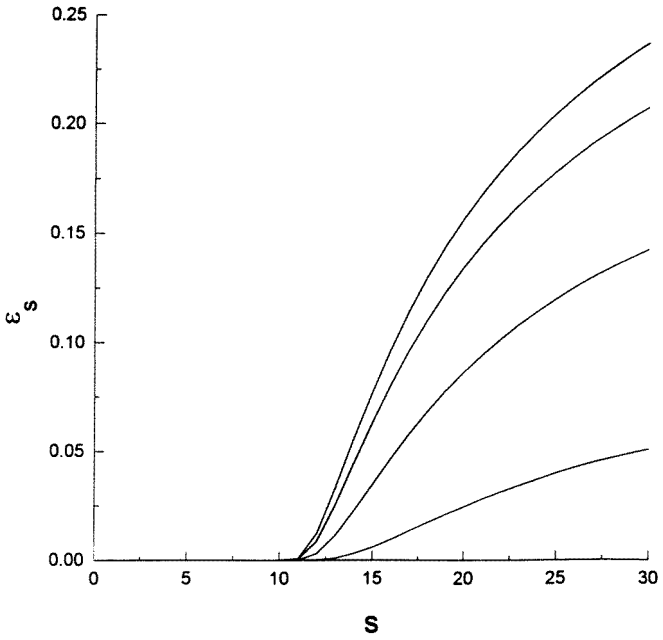

Figure 4. Average storage error as a function of the number of examples per category for $\alpha = 0.1$ and (from bottom to top) $d = 0.1$, 0.2, 0.3 and 0.4. The examples are stored without error for $s \leqslant 10$.

$$= \frac{1}{2}\sum_{m=0}^{s} B(m;s,b)\,\mathrm{erfc}\left[\frac{\overline{\gamma_s}(\eta)}{\sqrt{2\sigma_s^2(\eta)}}\right]. \tag{42}$$

In figure 4 we present the average storage error as a function of the number of examples per category s for fixed α and several values of the Hamming distance d. As expected, ϵ_c increases monotonically from $\epsilon_c = 0$ at $s = s_c$ to $\epsilon_c = \frac{1}{2}$ for $s \to \infty$.

5. Conclusion

The results presented in this paper indicate that the pseudo-inverse attractor neural network exhibits a categorization ability which is in all aspects similar to that found in the Hopfield model. This finding corroborates the view that the categorization property is a limitation of associative memory systems [19].

Some comments regarding the interpretation of the categorization error ϵ_c defined by (7) are in order. For the Hopfield model, the categorization error is defined by $(1-m^k)/2$ where m^k is the overlap between the concept $\boldsymbol{\xi}^k$ and the closest equilibrium state $\boldsymbol{S}_{\mathrm{eq}}$ [2]. Hence, it actually measures the average Hamming distance d_c between $\boldsymbol{\xi}^k$ and $\boldsymbol{S}_{\mathrm{eq}}$. Although the two definitions will coincide only if d_c equals the fraction of unstable sites in $\boldsymbol{\xi}^k$, which is given by (7), both quantities are equally good measures of the categorization ability provided that ϵ_c is a monotonically increasing function of d_c. This is a reasonable assumption which actually can be proved true in the regime of perfect storage [18]. We note that, in contrast to the Hopfield model, there is no analytical tool to characterize the equilibrium states of the pseudo-inverse beyond the storage capacity limit, so the calculation of d_c is not feasible at present.

Acknowledgments

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). CRN holds a FAPESP fellowship.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554
[2] Fontanari J F 1990 *J. Physique* **51** 2421
[3] Branchstein M C and Arenzon J J 1992 *J. Physique* I **2** 2019
[4] Krebs P R and Theumann W K 1993 *J. Phys. A: Math. Gen.* **26** 3983
[5] Silva C R, Tamarit F A, Lemke N, Arenzon J J and Curado E M 1995 *J. Phys. A: Math. Gen.* **28** 1593
[6] Kohonen T 1984 *Self-organisation and Associative Memory* (Berlin: Springer)
[7] Personnaz L, Guyon I and Dreyfus G 1986 *Phys. Rev.* A **34** 4217
[8] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[9] Opper M, Kleinz J, Köhler H and Kinzel W 1989 *J. Phys. A: Math. Gen.* **22** L407
[10] Henkel R D and Opper M 1991 *J. Phys. A: Math. Gen.* **24** 2201
[11] Gardner E 1989 *J. Phys. A: Math. Gen.* **22** 1969
[12] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
[13] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
[14] Fontanari J F 1993 *J. Phys. A: Math. Gen.* **26** 6147
[15] Binder K and Young A P 1986 *Rev. Mod. Phys.* **58** 801
[16] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[17] Kosterlitz J M, Thouless D J and Jones R C 1976 *Phys. Rev. Lett.* **36** 1217
[18] Rodrigues Neto C and Fontanari J F 1996 *J. Phys. A: Math. Gen.* **29** 3041
[19] Virasoro M A 1989 *Phys. Rep.* **184** 300