ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 61-66

Storage Capacity of the Reversed Wedge Perceptron with Binary Connections

G.J. Bex, R. Serneels

Limburgs Universitair Centrum
Universitaire Campus, B-3590 Diepenbeek, Belgium

Abstract. The storage capacity of a neural network with binary synaptic weights and a reversed wedge transfer function is studied both theoretically and numerically. Remarkably, for a particular width of the wedge, the storage capacity reaches a value equal to 1, which is the maximal value it can attain as follows from simple information theoretic considerations.

1. Introduction

In this paper we study the storage properties of a fully connected network with N^2 binary synaptic weights $\{J_{ij}\}$, $1 \leq i,j \leq N$, i.e. $J_{ij} = \pm 1$. The network's units S_i, $1 \leq i \leq N$ are restricted to values ± 1, and are updated according to

$$S_i(t+1) = \text{sgn}(f(\frac{1}{\sqrt{N}} \sum_{j=1}^{N} J_{ij} Sj(t))) \tag{1}$$

where $f(x)$ is an arbitrary function, and $g(x) = \text{sgn}(f(x))$ is called the transfer function.

The problem we will adress can be stated as follows: which is the maximum number $p_{stor.}$ of independent random patterns $\vec{\xi}^{\mu}$, $1 \leq \mu \leq p_{stor.}$ which are fixed points of the dynamics. These patterns are said to be stored by the network, and their number determines the storage capacity $\alpha = \frac{p_{stor.}}{N}$ of the system.

The study of the storage capacity of neural networks has progressed strongly in recent years. Storage capacities were studied for various types of networks. For the particular case of binary synaptic coefficients, and $f(x) = x$, leading to $g(x) = \text{sgn}(x)$ as a transfer function, which is implemented often in engineering applications, the storage capacity is known to be $\alpha \approx 0.83$ [1].

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 61-66

This is not the optimal value since the upper limit for the storage capacity is $\alpha = 1$, since one bit of information can be stored in each binary synaptic weight. The purpose of this contribution is to show that by changing the transfer function this maximum storage performance may be obtained. In particular we are thinking of a transfer function referred to as "reversed wedge" and described by $g(x) = \text{sgn}((x+K)x(x-K))$ where K is an arbitrary constant [2], see fig. a.

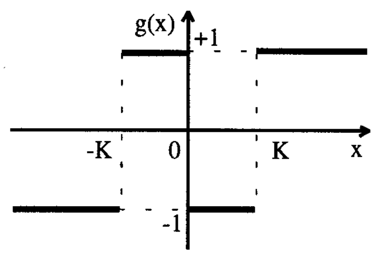

fig. a

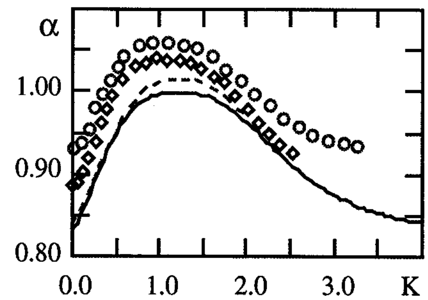

fig. b

Fig. a Reversed wedge transfer function
Fig. b Storage capacity vs. width of the wedge: some simulations ($N = 9 : \circ$; $N = 15 : \diamond$) and the extrapolation $N \longrightarrow \infty$ (dashed curve) are shown as well as the theoretic result (solid line)

For $K = 0$ the simple "$\text{sgn}(x)$" transfer function is recovered, while for $K \neq 0$ a wedge of width $2K$ has been inserted around $x = 0$. The idea is that by an adequate choice of K a neural network with a better storage capacity may be realised in the case of binary synapses. The storage capacity of a network with such a transfer function has already been studied by several authors in the case of continuous synaptic weights using the replica formalism of statistical mechanics [3, 4]. Calculations based on the replica symmetric Ansatz [3] show a highly increased storage capacity $\alpha \approx 10.5$ for $K \approx 1.18$, compared to the usual value of $\alpha = 2$ as determined by Gardner and Derrida [5]. As the validity of the replica symmetric assumption in this case is seriously in doubt; a one step replica symmetic calculation was made afterwards [4] and it was shown that the storage capacity could be at most $\alpha \approx 4.9$ for a value of $K \approx 0.8$.
In this contribution we perform in the case of binary synaptic coefficients a zero entropy replica symmetric calculation, which is generally assumed to be equivalent to a one step replica symmetry breaking calculation. Moreover, simulations are carried out to see if the quantitative result is correct.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 61-66

2. Zero entropy replica symmetric calculation

To study the storage capacity we consider a perceptron with N inputs in which we want to store p training patterns $\{\vec{\xi}^\mu, \xi_o^\mu\}$ for $\mu = 1 \dots p$. Using the by now standard statistical mechanical technique of Gardner and Derrida [5] we calculate the quantity

$$s = \frac{1}{N}\langle \log \Omega \rangle \tag{2}$$

which is called the entropy per synaps of the network. The brackets represent the average over the p random patterns chosen to be stored in the network and Ω is the number of synaptic coupling vectors which satisfy the p storage conditions. This number is given by:

$$\Omega = \sum_{\{\vec{J}\}} \prod_{\mu=1}^{p} \theta(g(\frac{\vec{J} \cdot \vec{\xi}^\mu}{\sqrt{N}})\xi_o^\mu)) \tag{3}$$

with $\theta(x)$ the usual Heaviside function. From this definition the value of Ω is a non negative integer. For given N, when the number of patterns p increases, the possibility to store all the patterns will decrease and so will Ω. As long as at least one set of synaptic coefficients will be found to store the patterns, Ω will be larger then one and the entropy will be positive. The critical storage capacity is reached when s becomes equal to zero.

A replica symmetric calculation [1, 6] yields the following expression for the entropy per synaps:

$$s = \underset{\{q,\hat{q}\}}{\mathrm{Extr}} \left(G_0(q,\hat{q}) - \alpha G_r(q)\right) \tag{4}$$

where $G_0(q,\hat{q})$ and $G_r(q)$ are given by:

$$G_0(q,\hat{q}) = -\frac{1}{2}(1-q)\hat{q} + \int_{-\infty}^{+\infty} dz \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \log\left(2\cosh(z\sqrt{\hat{q}})\right) \tag{5}$$

$$G_r(q) = \int_{-\infty}^{+\infty} dt\, \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \log\left(H(u_1) - H(u_2) + H(u_3)\right) \tag{6}$$

with the shorthand notations u_1, u_2, u_3:

$$u_1 = \frac{t\sqrt{q} - K}{\sqrt{1-q}} \; ; \; u_2 = \frac{t\sqrt{q}}{\sqrt{1-q}} \; ; \; u_3 = \frac{t\sqrt{q} + K}{\sqrt{1-q}} \tag{7}$$

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 61-66

and

$$H(x) = \int_x^\infty dt \, \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \tag{8}$$

The values of the order parameters q and $\hat{q}$ which extremise $G_0(q,\hat{q}) - \alpha G_r(q)$ in (4) are found from:

$$\hat{q} = \frac{\alpha}{2\pi(1-q)} \int_{-\infty}^{+\infty} dt \, \frac{e^{-\frac{1}{2}t^2}}{\sqrt{2\pi}} \left(\frac{e^{-\frac{1}{2}u_1^2} - e^{-\frac{1}{2}u_2^2} + e^{-\frac{1}{2}u_3^2}}{H(u_1) - H(u_2) + H(u_3)} \right)^2 \tag{9}$$

$$q = \int_{-\infty}^{+\infty} dz \, \frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}} \tanh^2(\sqrt{\hat{q}}z) \tag{10}$$

Substituting the solutions for q and $\hat{q}$ into (4) allows to trace out the entropy as a function of α and K. The storage capacity is found for the value of α for which the entropy is zero for fixed K. The so calculated storage capacity versus K is shown in fig. b. For $K = 0$ we find $\alpha \approx 0.83$ as required, for larger K values the storage capacity increases, attains a rather flat maximum of 1.0 for $K = \sqrt{2\log 2} \approx 1.18$ and decreases again towards $\alpha \approx 0.83$ for larger K values. So the theoretical treatment predicts an enhanced storage capacity for the reversed wedge model. It should be stressed that although the accuracy of this calculation depends on the replica symmetry Ansatz, the result is thought to be quite reliable. Due to the zero entropy condition this approach is known to be equivalent to a first step replica symmetry breaking calculation.

3. An upper bound for the storage capacity

The question for the replica formalism is the validity of the replica symmetric Ansatz, certainly for this type of problems. To a certain extent one may check this validity by independently calculating an upper bound for the storage capacity. This is done using the annealed approximation, i.e. in (2) not $\log\Omega$, but Ω is averaged over the patterns. Following [1, 6] the expression for the entropy per synaps in the annealed approximation is given by:

$$s_{ann.}(\alpha, K) = \frac{1}{N}\log\langle\Omega\rangle = \alpha\log\left(H(-K) - H(0) + H(K)\right) + \log 2 \tag{11}$$

Since $H(-K) = 1 - H(K)$ and $H(0) = \frac{1}{2}$ we find $s_{ann.}(\alpha, K) = (1-\alpha)\log 2$. As this annealed entropy is negative for all $\alpha > 1$, it follows immediately that storage capacities larger than 1.0 are in error. Actually

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 61-66

this result was expected from information theoretic arguments since in each binary synaptic weight only one bit of information can be stored [6]. This upper bound is clearly satisfied in our calculation as can be seen from the plot and so there is much confidence in these zero entropy replica symmetric calculations.

4. Simulations

To check further the theoretical results we performed simulations for a network with 9, 11, 13, 15 and 17 neurons. Although this is still far from the thermodynamic limit $N \longrightarrow \infty$ as assumed in the theoretical derivation, previous experience has shown that for networks of this size simulations give already the correct quantitative result [1]. The simulations were run using an enumeration technique [7], i.e. given the training examples we check numerically by enumeration of all possibilities whether a set of synaptic coefficients exists which stores the patterns. Because this enumeration method is time consuming, it is limited to a rather low number of neurons. Consequently the number of training examples which should be stored is slightly dependent on the particular choice of the patterns and we have to consider an average value of the storage capacity obtained after averaging over many choices of training examples (typically of the order of 10^4).
The resulting storage capacity as a function of the width of the wedge is shown in fig. b, we clearly observe a maximum storage capacity of $\alpha \approx 1$ at $K \approx 1.18$ which is in qualitative agreement with the replica symmetric calculation. The lack of precise agreement is thought to be a finite size effect, as for an increasing number of neurons the storage capacity decreases, and the extrapolated results for an infinite number of neurons are equal to the theoretical result up to a few percent. Furthermore the deviation from the theoretical result is systematic, which is explained by the correlations between the patterns for networks of this size. These tend to decrease the information content.

5. Conclusions

A replica symmetric calculation is performed for the storage capacity of a reversed wedge transfer function and an important increase of storage capacity is found. The stored information per synaps in this type of network is equal to the maximum set by information theory. This result is confirmed by simulations.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 61-66

References

[1] W. Krauth, M. Mézard: Storage capacity of memory with binary couplings. J. Physique, 50, 3057-3066 (1989)

[2] H. Nishimori, I. Opriş: Retrieval process of an associative memory with non-monotonic input-output function. IEEE Int. Conf. on Neural Networks, 1, 353-358 (1993)

[3] K. Kobayashi: On the capacity of a neuron with a non-monotonic output function. Network, 2, 237-243 (1991)

[4] G. Boffetta, R. Monasson, R. Zecchina: Symmetry breaking in non-monotonic neural networks. J. Phys. A: Math. Gen., 26, L507-513

[5] E. Gardner, B. Derrida: Optimal storage properties of neural networks. J. Phys. A: Math. Gen., 21, 271-284 (1988)

[6] M. Bouten, A. Komoda, R. Serneels: Storage capacity of a diluted neural network with Ising couplings. J. Phys. A: Math. Gen., 23, 2605-2612 (1990)

[7] W. Krauth, M. Opper: Critical storage capacity of the $J = \pm 1$ neural network. J. Phys. A: Math. Gen., 22, L519-523 (1989)