ANALYSIS OF THE HEBBIAN RULE
BY COUNTING METHODS

by Ph. PIRET

Philips Research Laboratory, Av. Albert Einstein 4, B-1348 Louvain-la-Neuve, Belgium

Abstract

Some properties of the Hebbian rule are analysed by means of counting methods. The emphasis is on the capacity and the error-correcting capability of the neural networks that implement this rule. This approach greatly simplifies the analysis of this class of networks.

Keywords: capacity, counting methods, Hebbian rule, Stirling approximation, union bound.

1. Introduction

The properties of the Hebbian rule have recently been investigated by several authors \(^1\text{–}^5\). To describe this rule we use the signature sgn(ξ) of a real number ξ: for ξ > 0 sgn(ξ) is +1, for ξ < 0 sgn(ξ) is −1, and for ξ = 0 sgn(ξ) is 0. We then extend the definition of the signature to matrices: if \(A_{ij}\) is the entry of a real matrix \(A\), sgn(\(A\)) is the matrix having sgn(\(A_{ij}\)) as the entry. Let \(U\) be the set \{+1, −1\} and let \(U_0\) be the set \{+1, 0, −1\}. To any \(m\times n\) matrix \(X\) over \(U\), one associates the \(n\times n\) matrix \(H = X^TX\) (with \(X^T\) the transpose of \(X\)) called the Hebbian matrix associated to \(X\). The Hebbian rule associated to \(X\) is then the mapping

\[ U_0^n \rightarrow U^n_0: u \mapsto h(u) = \text{sgn}(uH). \] (1)

It often preserves the rows of \(X\) and it has error-correcting capabilities: if \(u\) is “close” to the \(r\)th row \(X_r\) of \(X\), sgn(\(uH\)) is often equal to \(X_r \text{ }^2\text{–}^5\).

In this paper we show that many of these properties can be derived by simple counting arguments. Asymptotic results are then obtained by use of the Stirling approximation \(^6\text{,}^7\). Section 2 investigates the distribution of \((XH)_n\).
when $X$ goes through the set of all $m \times n$ matrices over $U$. This distribution is used in Sec. 3 to obtain bounds on the performances of the Hebbian rule. In Sec. 4, the error-correcting capabilities of the Hebbian rule are considered in more detail. In the sequel, $\log x$ denotes the natural logarithm of $x$ and $\mathcal{H}(x)$ denotes $-x \log x - (1-x) \log(1-x)$.

2. Distribution of $(XH)^s$ for arbitrary $X$

To obtain the value $(XH)^s = X_rX^TX^s$ (with $X^s$ the sth column of $X$) of an arbitrary element of $XH$ we decompose $X$ as $X = A + B$, where $B$ is obtained from $X$ by setting to zero all its elements in row $r$ and column $s$. It follows that $A$ is equal to the zero matrix except in row $r$ and column $s$ where it coincides with $X$, and that $X_rA^TX^s$ is equal to $(m+n-1)X^s_r$. Let $X$ now go through the set $S(X_r, X^s)$ of all $m \times n$ matrices over $U$ with prescribed values, $X_r$ and $X^s$, of their rth row and sth column. In this case $A$ is fixed and $B$ goes through the set of all $m \times 1$ matrices satisfying $B_r = 0$ and $B^s = 0$. The quantity $X_rB^TX^s$, is then easily seen to be a sum of $(m-1)(n-1)$ independent elements of $U$. As a result, for any $j$, there are exactly

$$\binom{(m-1)(n-1)}{j}$$

matrices $B$ for which this sum contains $j$ terms equal to $-1$ and $(m-1)(n-1) - j$ terms equal to $+1$. Equivalently, the possible values of $X_rB^TX^s$ are those $w$ of the same parity as $(m-1)(n-1)$ satisfying $|w| \leq (m-1)(n-1)$, and for any such $w$, the fraction of matrices $B$ achieving $X_rB^TX^s = w$ is given by

$$\phi(w) = \frac{(m-1)(n-1)}{2[(m-1)(n-1) - w]} 2^{-(m-1)(n-1)}.$$  \hfill (2)

Summing the contributions of $A$ and $B$ we have

**Theorem 1**: For any $r$, $s$, $X_r$ and $X^s$, the fraction of matrices $X$ in $S(X_r, X^s)$ that satisfy

$$(XH)^s = (m + n - 1 + w)X^s_r$$

is equal to $\phi(w)$. 

---

Ph. Piret
Corollary 2: The mean value $\sum_w (m + n - 1 + w) \phi(w)$ of $(XH)^r_\ell$ for $X \in S(X_r, X^\delta)$, is equal to $(m + n - 1)X^\delta_r$.

These two statements remain obviously true when $X$ goes through the larger sets $S(X_r)\text{ or } S(X^\delta_r)$ of all $m \times n$ matrices over $U$ having a prescribed value of $X_r$ or $X^\delta_r$.

Let us now compute the fraction $F^\delta_r(\lambda)$ of all $m \times n$ matrices $X \in S(X_r, X^\delta)$ that violate at least one of the conditions

\[ \text{sgn}(XH)^r_\ell = X^\delta_r, \quad |(XH)^r_\ell| > \lambda(m + n - 1), \]  

for some real $\lambda \in [0, 1]$. This fraction is given by

\[ F^\delta_r(\lambda) = \sum_w \phi(w), \]  

where the sum is over all $w$ having the same parity as $(m-1)(n-1)$ and satisfying

\[ w \leq -(1 - \lambda)(m + n - 1). \]

Since $\phi(w)$ is symmetric with respect to $w = 0$, the last inequality can be replaced by

\[ w \geq (1 - \lambda)(m + n - 1). \]

Applying to (4) the well-known formula

\[ \sum_{i=1}^{b} \binom{b}{i} \leq \frac{l}{2l-b} \binom{b}{1} \text{ for all } l > \frac{b}{2}, \]

we obtain

\[ F^\delta_r(\lambda) \leq \frac{mn - \lambda(m + n - 1)}{2(1 - \lambda)(m + n - 1)} \binom{(m - 1)(n - 1)}{k} 2^{-(m-1)(n-1)}, \]

where $k$ is the largest integer less than or equal to $[mn - \lambda(m + n - 1)]/2$. In (5) we still need to evaluate the binomial coefficient, which has the form

\[ \binom{b}{b(1 + \omega)/2} \]
with

\[ b = (m - 1)(n - 1) \]

and

\[ \omega = (1 - \lambda)(m + n - 1)/(m - 1)(n - 1). \]

The use of the Stirling approximation leads to

\[
\left( \frac{b}{b(1 + \omega)/2} \right) \leq \left[ \frac{2}{\pi b(1 - \omega^2)} \right]^{1/2} \exp \left[ b, \mathcal{H} \left( \frac{1 + \omega}{2} \right) \right].
\]

(6)

Then using \( \mathcal{H}[(1 + \omega)/2] \leq \log 2 - \omega^2/2 \), we obtain

\[
\left( \frac{b}{b(1 + \omega)/2} \right) \leq \left[ \frac{2}{\pi b(1 - \omega^2)} \right]^{1/2} 2^b \exp \left( -\frac{b\omega^2}{2} \right).
\]

(7)

Finally, the use of (7) in (5), with the appropriate value of \( b \), leads to

\[
F^*_x(\lambda) \leq \frac{mn - \lambda(m + n - 1)}{2(1 - \lambda)(m + n - 1)} \left[ \frac{2}{\pi(m - 1)(n - 1)(1 - \omega^2)} \right]^{1/2} \exp[-J(m, n, \lambda)],
\]

(8)

where \( J \) is given by

\[
J(m, n, \lambda) = \frac{(1 - \lambda)^2(m + n - 1)^2}{2(m - 1)(n - 1)}.
\]

The quantity \( F^*_x(\lambda) \) was defined above to be the fraction of matrices \( X \in S(X^r, X^s) \) that do not satisfy (3). Obviously, the remarks made after corollary 2 imply that it also represents the fraction of \( X \in S(X^r) \) that do not satisfy (3). In the sequel, (8) is used to obtain bounds on the performances of the rule (1).

3. Bounds on the performances of \( X \mapsto \text{sgn}(XH) \)

First, we consider the case where \( m \) is proportional to \( n: m = [\alpha n] \) for some real \( \alpha > 0 \). In this case, we define \( \phi(\alpha, \lambda) \) to be the limit of the right-hand side of (8) for \( n \to \infty \). One easily obtains

\[
\phi(\alpha, \lambda) = \frac{1}{(1 - \lambda)(1 + \alpha)} \left( \frac{\alpha}{2\pi} \right)^{1/2} \exp \left[ -\frac{(1 - \lambda)^2(1 + \alpha)^2}{2\alpha} \right].
\]

(9)
Analysis of the Hebbian rule by counting methods

Of particular interest is the value

\[ \phi(\alpha, 0) = \frac{1}{1 + \alpha} \left( \frac{\alpha}{2\pi} \right)^{1/2} \exp \left[ -\frac{(1 + \alpha)^2}{2\alpha} \right], \]  

(10)
giving an upper bound, for \( n \to \infty \), on the fraction of \([\alpha n] \times n\) matrices \( X \) that do not satisfy \( \text{sgn}(XH)^s_r = X^s_r \). The maximum value of \( \phi(\alpha, 0) \) is obtained for \( \alpha = 1 \):

\[ \phi(1, 0) = e^{-2}/\sqrt{8\pi} \approx 0.0270. \]  

(11)

It may be a surprise to see that \( \lim_{\alpha \to \infty} \phi(\alpha, 0) \) is zero. However, let us remark that when \( X \) is the "complete" \( 2^n \times n \) matrix having as rows the \( 2^n \) different elements of \( U^n \), then \( H \) is \( 2^n \) times the \( n \times n \) identity matrix. In this case, all rows of \( X \) are preserved by \( X \leftrightarrow \text{sgn}(XH) \). This agrees with \( \lim_{\alpha \to \infty} \phi(\alpha, 0) = 0 \).

We consider now the case where \( m - 1 \) is proportional to \((n - 1)/\log(n - 1)\)

\[ (m - 1) = \lfloor \beta(n - 1)/\log(n - 1) \rfloor, \quad \beta > 0, \]  

(12)
and we define \( \psi(\beta, \lambda) \) to be the right-hand side of (8) under (12). For large \( n \), one easily obtains

\[ F_r(\lambda) \leq \psi(\beta, \lambda) \leq \left[ \frac{\beta}{2(1 - \lambda)^2 \pi \log(n - 1)} \right]^{1/2} (n - 1)^{-(1 - \lambda)^2/2\beta + \sigma(1)}. \]  

(13)

Let \( S(X_r) \) be the set of \( m \times n \) matrices \( X \) having \( X_r \) as \( r \)th row and let \( F_r(\lambda) \) be the proportion of \( X \in S(X_r) \) that do not satisfy (3) for at least one \( s \). Let also \( F(\lambda) \) be the proportion of matrices \( X \) that do not satisfy (3) for at least one pair \((r, s)\). The remarks made after corollary 2 and at the end of Sec. 2 together with the well-known union bound imply \( F_r(\lambda) \leq nF^s_r(\lambda) \) and \( F(\lambda) \leq mnF^s_r(\lambda) \), which means, by use of (12) and (13),

\[ F_r(\lambda) \leq \left[ \frac{\beta}{2(1 - \lambda)^2 \pi \log(n - 1)} \right]^{1/2} (n - 1)^{1 - (1 - \lambda)^2/2\beta + \sigma(1)}, \]  

(14)

\[ F(\lambda) \leq \left[ \frac{\beta^3}{2(1 - \lambda)^2 \pi \log(n - 1)^3} \right]^{1/2} (n - 1)^{2 - (1 - \lambda)^2/2\beta + \sigma(1)}. \]  

(15)

Equations (13), (14) and (15) have the following interpretation.
Theorem 3: For \( m \) satisfying (12) and for sufficiently large \( n \), the fraction \( F'_n(\lambda) \) goes to zero for any \( \beta \geq 0 \), the fraction \( F_n(\lambda) \) goes to zero for \( \beta < (1 - \lambda)^2/2 \) and the fraction \( F(\lambda) \) goes to zero for \( \beta < (1 - \lambda)^2/4 \).

These statements are due to McEliece et al.\(^2\). Another proof is given by Kuh and Dickinson \(^3\). We think that the counting proof given above is the simplest one to date.

4. Error correction by means of \( y \mapsto \text{sgn}(yH) \)

In this section we always assume that \( m \) satisfies (12) for some positive value of \( \beta \). Then let \( X_r \) and \( y \) be two arbitrary elements of \( U^n \) satisfying \( d(X_r, y) = t < n/2 \), with \( d(X_r, y) \) the Hamming distance between \( X_r \) and \( y \). We first investigate which is the fraction of \( m \times n \) matrices \( X \) having \( X_r \) as \( r \)th row and satisfying \( \text{sgn}(yH) = X_r \) for \( H = X^T X \). Then we show that for \( \beta \) small enough and for most \( m \times n \) matrices \( X \) over \( U \), most \( y \in U^n \) at distance \( \leq t(\beta) < n/2 \) of an arbitrary row \( X_r \) of \( X \) satisfy \( \text{sgn}(yH) = X_r \).

Let us begin with the first of these problems. As in Sec. 2, we write \( yH^s = yA^T X^s + yB^T X^s \). For those \( s \) satisfying \( y^s = X_r^s \), \( yA^T X^s \) is equal to \( (n + m - 1 - 2t)X_r^s \) and for those \( s \) satisfying \( y^s = -X_r^s \), \( yA^T X^s \) is equal to \( (n - m + 1 - 2t)X_r^s \). As for the second term \( yB^T X^s \), it is equal to \( w \) for a fraction \( \phi(w) \) of the \( 2^{m-1}(n-1) \) possible values of \( B \).

These remarks and the well-known union bound imply that the fraction \( Q(t) \) of matrices \( X \) containing a specified row \( X_r \) at distance \( t \) of some arbitrary \( y \) and achieving \( \text{sgn}(yH) \neq X_r \) satisfies

\[
Q(t) \leq t \sum_{w \in 2^{t-1} - n + m} \phi(w) + (n - t) \sum_{w \in 2^{t+1} - n - m} \phi(w). \tag{16}
\]

By the same techniques as those leading to (8) the first of these sums is shown to be at most \( \gamma_1 \gamma_2 \gamma_3 \), with

\[
\gamma_1 = \frac{mn - 2m - 2t + 2}{2(n + 1 - m - 2t)}, \quad \gamma_2 = \left[ \frac{2}{\pi(m - 1)(n - 1)(1 - \gamma)^2} \right]^{1/2},
\]

\[
\gamma_3 = \exp \left[ -\frac{(n + 1 - m - 2t)^2}{2(m - 1)(n - 1)} \right], \quad \gamma = \frac{n + 1 - m - 2t}{(m - 1)(n - 1)}.
\]

while the second one is at most \( \delta_1 \delta_2 \delta_3 \), with

\[
\delta_1 = \frac{mn - 2t}{2(n - 1 + m - 2t)}, \quad \delta_2 = \left[ \frac{2}{\pi(m - 1)(n - 1)(1 - \delta)^2} \right]^{1/2},
\]

438 Philips Journal of Research Vol. 45 No. 6 1991
Analysis of the Hebbian rule by counting methods

For \( n \to \infty, \tau = [nt] \) with \( 0 < \tau < 1/2 \), and \( m \) satisfying (12), we see that \( \gamma_1 \) and \( \delta_1 \) are approximately \( \beta(n - 1)/2(1 - 2\tau) \log(n - 1) \), \( \gamma_2 \) and \( \delta_2 \) are approximately \( n^{-1}[2 \log(n - 1)/\pi \beta]^{1/2} \), \( \gamma_3 \) and \( \delta_3 \) are approximately \( (n - 1)^{-1} - (1 - 2\tau)^2/2 \beta \) and \( \gamma \) and \( \delta \) are approximately zero. Writing \( Q(\tau) \) in place of \( Q([nt]) \) we thus obtain

\[
Q(\tau) \leq \left[ \frac{\beta}{2\pi (1 - 2\tau)^2 \log(n - 1)} \right]^{1/2} (n - 1)^{1 - (1 - 2\tau)^2/2 \beta + o(1)} \tag{17}
\]

which leads to

**Theorem 4:** Assume that \( m \) satisfies (12) and choose \( \tau < 1/2 \). For given \( X_r \) and \( y \in U^n \) satisfying \( d(X_r, y) = [nt] \), the fraction of \( m \times n \) matrices \( X \) having \( X_r \) as \( r \)th row and satisfying \( \text{sgn}(yX^T) \neq X_r \) goes to zero for \( n \to \infty \) when \( \beta \) is less than \((1 - 2\tau)^2/2\).

Theorem 4 implies that when \( X_r \) and \( y \) satisfy \( d(X_r, y) = [nt] \) with \( \tau < 1/2 \) and when \( m \) is given by (12), most \( m \times n \) matrices containing \( X_r \) as a row do not contain another row at a distance less than \( n/2 \) of \( y \). This is a well-known result of coding theory \(^6\).

Let us now investigate for how many \( m \times n \) matrices \( X \) over \( U \) containing an arbitrary \( X_r \) as the \( r \)th row, most \( y \in U^n \) at the Hamming distance \( t = [nt] < n/2 \) of \( X_r \) satisfy \( \text{sgn}(yH) = X_r \). We denote first by \( \mathcal{H}(t) \) the set of all \( u \in U_n \) having \( (n - t) \) entries equal to \( +1 \) and \( t \) entries equal to \( -1 \). Then we construct an array \( \mathcal{A} \) with \( (\mathcal{H}(t)) \) rows and \( 2(m - 1)n \) columns. Its rows are indexed by the elements \( u \) of \( \mathcal{H}(t) \) and its columns are indexed by the \( m \times n \) matrices \( X \) having a prescribed value \( X_r \) of their \( r \)th row. For given \( u \) and \( X_r \) in \( U^n \), let \( u \ast X_r \) denote the componentwise product of \( u \) and \( X_r \) : \( (u \ast X_r)^i = u^i X_r^i \). The set \( \{ u \ast X_r : u \in \mathcal{H}(t) \} \) is then nothing but the set of \( y \in U^n \) satisfying \( d(y, X_r) = t \). To fill in the array, we write a zero in position \( (u, X_r) \) when \( \text{sgn}[(u \ast X_r)H] \) is equal to \( X_r \). Otherwise we write a one in this position.

Assume now that \( \beta \) is less than \((1 - 2\tau)^2/2\), choose \( \epsilon > 0 \) satisfying \( 1 + \epsilon - (1 - 2\tau)^2/2 \beta < 0 \) and define \( Q_\epsilon(\tau) = (n - 1)^\epsilon Q(\tau) \). We know from the arguments leading to theorem 4 that the rows of \( \mathcal{A} \) contain \( 2^{(n - 1)n}Q(\tau) \) ones, which implies that the total number of ones in \( \mathcal{A} \) is \( (\mathcal{H}(t))2^{(n - 1)n}Q(\tau) \). Hence, if we denote by \( M(\epsilon) \) the number of columns containing \( (\mathcal{H}(t))Q_\epsilon(\tau) \) ones or more, we obtain
which is equivalent to

\[ M(\varepsilon)/2^{m-1}n < (n - 1)^{-\varepsilon}, \]

and leads to the following theorem.

**Theorem 5:** Assume that \( m \) satisfies (12) with \( \beta < (1 - 2\tau)^2/2 \) and \( 0 \leq \tau < 1/2 \). In this case, for \( n \) large enough, most \( m \times n \) matrices \( X \) over \( U \) with a prescribed row \( X_r \) are such that most \( n \)-tuples at distance \( \leq \lfloor \eta \tau \rfloor \) of \( X_r \) satisfy \( \text{sgn}(yH) = X_r \).

**Proof:** Let \( \alpha(n) \) be the proportion of matrices \( X \) with prescribed \( X_r \) for which a fraction strictly larger than

\[ \eta(n) = (n - 1)^{1 + \varepsilon - (1 - 2\tau)^2/2\beta + o(1)} \]

of the \( n \)-tuples \( u \ast X_r, \ u \in \mathcal{U}(t) \), does not satisfy \( \text{sgn}[(u \ast X_r)H] = X_r \). It follows from (18) that \( \lim_{n \to \infty} \alpha(n) = 0 \). Since

\[ 1 + \varepsilon - (1 - 2\tau)^2/2\beta < 0, \]

\( \lim_{n \to \infty} \eta(n) \) is also zero and the theorem is proved.

Let us construct another array with \( \binom{n}{\tau} \) rows and \( 2^{m^2} \) columns. The rows of this array are still indexed by the elements \( u \) of \( \mathcal{U}(t) \) and its columns are now indexed by the \( m \times n \) matrices over \( U \). In position \( (u, X) \) the entry of the array is zero if \( \text{sgn}[(u \ast X_i)H] = X_i \) is satisfied for \( 1 \leq i \leq m \). Otherwise, the \( (u, X) \) entry is one. Using the union bound, we see that the fraction \( Q^*(\tau) \) of ones in any row of this new array is at most \( mQ(\tau) \). By use of (12) and (17) this implies

\[ Q^*(\tau) \leq \left\{ \frac{\beta^3}{2\pi(1 - 2\tau)^2[\log(n - 1)]^3} \right\}^{1/2} (n - 1)^{2 - (1 - 2\tau)^2/2\beta + o(1)}. \]  

(19)

For \( \beta < (1 - 2\tau)^2/4 \), choose \( \varepsilon > 0 \) satisfying \( 2 + \varepsilon - (1 - 2\tau)^2/2\beta < 0 \), define \( Q^*_\varepsilon(\tau) = (n - 1)^{\varepsilon}Q^*(\tau) \) and denote by \( M^*(\varepsilon) \) the number of columns of the array containing \( \binom{n}{\tau}Q^*_\varepsilon(\tau) \) or more ones. One obtains

\[ M^*(\varepsilon) \binom{n}{\tau}Q^*_\varepsilon(\tau) \leq 2^{mn}Q^*(\tau) \binom{n}{\tau}, \]
Analysis of the Hebbian rule by counting methods

which is equivalent to

$$M^*(e)/2^m < (n - 1)^{-e},$$  \hspace{1cm} (20)$$

and leads to the following theorem.

**Theorem 6:** Assume that $m$ satisfies (12) with $\beta < (1 - 2\tau)^2/4$ and $0 \leq \tau < 1/2$. In this case, for $n$ large enough, most $m \times n$ matrices $X$ over $U$ satisfy

$$\text{sgn}(u \ast X_i) = X_i, \quad 1 \leq i \leq m,$$

for most $u \in \mathcal{U}(\lfloor n\tau \rfloor)$.

**Proof:** Let $\alpha^*(n)$ be the proportion of matrices $X$ for which a fraction strictly larger than

$$\eta^*(n) = (n - 1)^2 + \varepsilon - (1 - 2\tau)^2/2\beta + o(1)$$

of the $n$-tuples $u \in \mathcal{U}(t)$ does not satisfy $\text{sgn}[(u \ast X_i)H] = X_i$ for $1 \leq i \leq m$. It follows from (20) that $\lim_{n \to \infty} \alpha^*(n)$ is zero. Since

$$2 + \varepsilon - (1 - 2\tau)^2/2\beta < 0,$$

$\lim_{n \to \infty} \eta^*(n)$ is also zero and the theorem is proved. \hspace{1cm} $\square$

5. Conclusion

We have used counting arguments to derive several properties of the Hebbian rule. These arguments are quite simple and easily lead to asymptotic expressions when coupled with Stirling's approximation. We feel that this approximation, when compared with other methods of proof, is more efficient to obtain asymptotic expressions for problems of this type.

In this paper we did not analyse the exact situation considered in other references $^2,^3$) where the matrix $H = X^TX$ was replaced by $H^* = X^TX + (k - m)I_n$, for some $k$ in $[0, m]$. The main consequence of considering $H^*$ in place of $H$ is that $m$ should be replaced by $k$ at several places. For $m$ satisfying (12) this does not modify the asymptotic results. For $m = \lfloor \alpha n \rfloor$ and $k = \lfloor \gamma m \rfloor$ with $0 < \gamma < 1$, $\alpha$ should be replaced by $\alpha \gamma$ in (9).

REFERENCES


Author