NOTE **Communicated by Sara Solla**

A Simple Derivation of a Bound on the Perceptron Margin Using Singular Value Decomposition

Omri Barak
ob2194@columbia.edu
Center for Theoretical Neuroscience, Department of Neuroscience, Columbia University Medical Center, New York, NY 10032, U.S.A.

Mattia Rigotti
mr2666@columbia.edu
Center for Theoretical Neuroscience, Department of Neuroscience, Columbia University Medical Center, New York, NY 10032, U.S.A., and Center for Neural Science, New York University, NY 10003, U.S.A.

The perceptron is a simple supervised algorithm to train a linear classifier that has been analyzed and used extensively. The classifier separates the data into two groups using a decision hyperplane, with the margin between the data and the hyperplane determining the classifier's ability to generalize and its robustness to input noise. Exact results for the maximal size of the separating margin are known for specific input distributions, and bounds exist for arbitrary distributions, but both rely on lengthy statistical mechanics calculations carried out in the limit of infinite input size. Here we present a short analysis of perceptron classification using singular value decomposition. We provide a simple derivation of a lower bound on the margin and an explicit formula for the perceptron weights that converges to the optimal result for large separating margins.

1 Introduction

The perceptron is a simple algorithm for training a linear classifier to separate a data set into two distinct classes (Rosenblatt, 1962). It works by iteratively updating a weight vector to define a decision hyperplane that separates the inputs into the two desired classes (Minsky & Papert, 1969). In addition to its simplicity, the perceptron algorithm has the appealing property of converging after a finite number of iterations if the data set is linearly separable (Novikoff, 1962).

More recent modifications of the original perceptron have led to algorithms that are guaranteed to converge to an optimal solution—one corresponding to a decision hyperplane that maximally separates the two data classes. This is obtained by maximizing the separating margin, defined as the distance between the input classes and the decision hyperplane (Krauth

& Mézard, 1987; Freund & Schapire, 1999; Korzen & Klesk, 2008; Abbott & Kepler, 1989a). This strategy increases the classifier's robustness to input noise and its ability to generalize to untrained data.

In her seminal work, Gardner (1988) proved an exact relation between the number of patterns the perceptron has to classify and the maximal margin attainable. This result, however, holds only in the thermodynamical limit (where the number of neurons and input patterns goes to infinity) and for independent and identically distributed inputs. Bounds on the margin were later obtained by Tarkowski, Komarnicki, and Lewenstein (1991) and Tarkowski and Lewenstein (1992) for a general distribution of inputs through a replica method analysis (Mézard, Parisi, & Virasoro, 1987).

Our main result is an independent derivation of the bound obtained by Tarkowski et al. (1991) and Tarkowski and Lewenstein (1992) using elementary linear algebra methods, including singular value decomposition. Specifically, we show that the margin is bounded by the minimal singular value of the matrix whose columns are the input patterns. This result is valid for any set of input patterns and does not assume any particular correlation structure. Our analysis also provides a straightforward derivation of the pseudo-inverse solution to the perceptron (Personnaz, Guyon, & Dreyfus, 1985; Kanter & Sompolinsky, 1987), which provides a closed-form expression for the weights of the perceptron that converges to the optimal solution for large values of the separating margin.

The perceptron has been used as a tool in a variety of fields ranging from machine learning (Freund & Schapire, 1999), through modeling of specific brain regions (Brunel, Hakim, Isope, Nadal, & Barbour, 2004) to training methods for spiking and decision-making neural networks (Brader, Senn, & Fusi, 2007; Rigotti et al., 2010). A simple way of analyzing the perceptron can provide valuable insight into all these fields.

2 Framework

We consider a perceptron with N binary inputs and a single output. The perceptron has to separate p patterns $\xi_i^\mu = \pm 1$ into two classes $\zeta^\mu = \pm 1$, where $i = 1, \ldots, N$ and $\mu = 1, \ldots, p$. The ratio between the number of patterns and the number of inputs defines the storage capacity $\alpha = p/N$. The output of the perceptron is determined by its weights w_i and, for a given pattern $\boldsymbol{\xi}^\mu$, is defined as $o^\mu = \text{sign}\left(\sum_i w_i \xi_i^\mu\right)$. Therefore, the conditions for correct classification are

$$h^\mu = \zeta^\mu \sum_i^N w_i \xi_i^\mu \geq \kappa, \qquad \mu = 1, \ldots, p \tag{2.1}$$

with $\kappa > 0$.

To ensure robustness to input noise and allow generalization, it is useful to maximize κ, with the constraint that the vector of weights w_i lies on the unit sphere (see Gardner, 1988). A solution to the full problem is therefore a weight vector $\mathbf{w}$ satisfying the conditions

$$\begin{aligned} \sum_i^N w_i^2 &= 1 \\ h^\mu &\geq \kappa, \end{aligned} \tag{2.2}$$

for all μ and for a given $\kappa > 0$. The optimal solution maximizes κ.

To simplify the presentation of the analysis, we use matrix notation. We define an $N \times p$ matrix S with components $S_{i\mu} = \zeta^\mu \xi_i^\mu$. Equations 2.1 and 2.2 then read:

$$\begin{aligned} h^\mu &\geq \kappa \\ \mathbf{h}^T &= \mathbf{w}^T S \\ \mathbf{w}^T \mathbf{w} &= 1, \end{aligned} \tag{2.3}$$

for all μ and a given $\kappa > 0$, and where $\mathbf{w}^T$ and $\mathbf{h}^T$ denote the row vectors obtained by transposing $\mathbf{w}$ and $\mathbf{h}$.

We now factorize the input matrix using singular value decomposition (SVD),

$$S = U \Sigma V^T, \tag{2.4}$$

where (defining r as the rank of S) U is an $N \times r$ matrix with orthonormal columns ($U^T U = I$), Σ is an $r \times r$ diagonal matrix with positive real numbers on the diagonal (the singular values of S) and V is a $p \times r$ matrix with orthonormal columns ($V^T V = I$).

The SVD decomposition of the input matrix S suggests an equivalent perceptron problem obtained by absorbing the matrix U into the weight vector $\mathbf{w}$. The original and equivalent formulations are:

- For $\kappa > 0$, find $\mathbf{w}$ with $\mathbf{w}^T \mathbf{w} = 1$, so that $h^\mu \geq \kappa$, for all μ, where $\mathbf{h}^T = \mathbf{w}^T (U \Sigma V^T)$, (P1)
- For $\widetilde{\kappa} > 0$, find $\widetilde{\mathbf{w}}$ with $\widetilde{\mathbf{w}}^T \widetilde{\mathbf{w}} = 1$, so that $\widetilde{h}^\mu \geq \widetilde{\kappa}$, for all μ, where $\widetilde{\mathbf{h}}^T = \widetilde{\mathbf{w}}^T (\Sigma V^T)$. (P2)

The first formulation $P1$ is a restatement of the original problem. We will show that the second form $P2$ is equivalent to the original formulation, in that there exists a transformation from the optimal solution to $P2$ to the optimal solution to $P1$, and vice versa. Working with the

formulation $P2$ will allow us to derive a lower bound on $\widetilde{\kappa}$ and, in turn, on κ.

3 Equivalence of $P1$ and $P2$: Uncovering the True Dimensionality of the Problem

We now show that if a weight vector $\widetilde{\mathbf{w}}$ is the optimal solution of $P2$, then $\mathbf{w} = U\widetilde{\mathbf{w}}$ is the optimal solution of $P1$. This is equivalent to saying that if $\widetilde{\mathbf{w}}$ satisfies $P2$ with $\widetilde{\kappa}$, then $\mathbf{w}$ satisfies $P1$ with $\kappa = \widetilde{\kappa}$. The statement follows from the definitions $P1$ and $P2$ and from the normalization of $\mathbf{w}$:

$$\mathbf{w}^T\mathbf{w} = \widetilde{\mathbf{w}}^T U^T U\widetilde{\mathbf{w}} = 1.$$

Now we prove by contradiction that $\mathbf{w} = U\widetilde{\mathbf{w}}$ is actually the optimal solution to $P1$ if $\widetilde{\mathbf{w}}$ is the optimal solution to $P2$. Suppose there exists a better solution than $\mathbf{w}$ to $P1$, that is, there is a normalized weight vector $\mathbf{q}$ that satisfies

$$\min(\mathbf{q}^T S) = \kappa' > \kappa = \min(\mathbf{w}^T S).$$

If we now define the normalized r-dimensional vector $\widetilde{\mathbf{q}} = U^T\mathbf{q}/(\mathbf{q}^T UU^T\mathbf{q})$ as a candidate $P2$ solution, we can see that it satisfies

$$\min\left(\widetilde{\mathbf{q}}^T \Sigma V^T\right) = \frac{\min\left(\mathbf{q}^T U\Sigma V^T\right)}{\left\|U^T\mathbf{q}\right\|^2} = \frac{\kappa'}{\left\|U^T\mathbf{q}\right\|^2} \geq \kappa' > \kappa = \widetilde{\kappa},$$

contradicting the optimality of $\widetilde{\mathbf{w}}$. The inequality $\|U^T\mathbf{q}\|^2 \leq \|\mathbf{q}\|^2$ stems from the fact that U is a projection on an r-dimensional vector space.

Notice that in general, the matrix ΣV^T of the perceptron problem, $P2$, is an $r \times p$ matrix, where r is the rank of S. $P1$ and $P2$ are therefore equivalent to an r-dimensional perceptron classifying p patterns. In particular, the reformulation, $P2$, uncovers the true dimensionality of the classification problem at hand.

4 Dual Formulation and Lower Bound on κ

Assuming that a solution to $P2$ exists for some $\widetilde{\kappa} > 0$, we now reformulate the task of finding the optimal solution to the perceptron problem—the maximal $\widetilde{\kappa}$ for which $P2$ holds. Defining $\bar{\mathbf{h}} = \widetilde{\mathbf{h}}/\widetilde{\kappa}$ and $\bar{\mathbf{w}} = \widetilde{\mathbf{w}}/\widetilde{\kappa}$ and substituting these definitions in the equivalent equation, $P2$, allows us to reformulate the search for the optimal solution as the problem of finding a weight vector

$\bar{\mathbf{w}}$ satisfying

$$\bar{\mathbf{w}}^T \bar{\mathbf{w}} = \frac{1}{\widetilde{\kappa}^2} \quad \text{subject to } \bar{\mathbf{h}} \geq 1, \quad \text{where } \bar{\mathbf{h}}^T = \bar{\mathbf{w}}^T \Sigma V^T,$$

for the largest possible $\widetilde{\kappa} > 0$, where the notation $\bar{\mathbf{h}} \geq 1$ stands for $\min_\mu \left(\bar{h}^\mu\right) \geq 1$. Because maximizing $\widetilde{\kappa} > 0$ is the same as minimizing $1/\widetilde{\kappa}^2$, our task can be formulated in the following equivalent dual form: instead of maximizing $\min_\mu \left(\bar{h}^\mu\right)$ subject to a constraining equality on $\bar{\mathbf{w}}^2$, we minimize $\bar{\mathbf{w}}^2$ subject to inequality constraints on $\bar{h}^\mu$:

$$\mathbf{w}^* = \arg\min_{\bar{\mathbf{w}}} \left(\bar{\mathbf{w}}^T \bar{\mathbf{w}}\right) \quad \text{subject to } \bar{\mathbf{h}} \geq 1, \quad \text{where } \bar{\mathbf{h}}^T = \bar{\mathbf{w}}^T \Sigma V^T.$$

The maximal margin κ^* will then be given by

$$1/\kappa^{*2} = \mathbf{w}^{*T} \mathbf{w}^* = \min_{\bar{\mathbf{w}}} \left(\bar{\mathbf{w}}^T \bar{\mathbf{w}}\right) = \min_{\bar{\mathbf{w}}} \left(\bar{\mathbf{h}}^T M \bar{\mathbf{h}}\right) \tag{4.1}$$
$$\text{subject to } \bar{\mathbf{h}} \geq 1, \quad \text{where } \bar{\mathbf{h}}^T = \bar{\mathbf{w}}^T \Sigma V^T,$$

and we define a $p \times p$ matrix $M = V \Sigma^{-2} V^T$.

We now consider the special case $r = p$ (which implies that there are $p \leq N$ linearly independent patterns). In this case, the formula $\bar{\mathbf{w}} = \Sigma^{-1} V^T \bar{\mathbf{h}}$ defines a one-to-one relationship between $\bar{\mathbf{h}}$ and $\bar{\mathbf{w}}$. Thus, we can switch the minimization parameter from $\bar{\mathbf{w}}$ to $\bar{\mathbf{h}}$:

$$1/\kappa^{*2} = \min_{\bar{\mathbf{h}}:\ \bar{\mathbf{h}} \geq 1} \bar{\mathbf{h}}^T M \bar{\mathbf{h}}. \tag{4.2}$$

This expression is equivalent to equation 24 in Tarkowski et al. (1991). We now use this result to obtain a lower bound on the maximal margin κ^*. Specifically, observe that the all-ones vector $\mathbf{1}^T = (1, 1, ..., 1)$ satisfies the condition $\bar{h}^\mu \geq 1$. Using this fact in equation 4.2, we get

$$1/\kappa^{*2} = \min_{\bar{\mathbf{h}}:\ \bar{\mathbf{h}} \geq 1} \bar{\mathbf{h}}^T M \bar{\mathbf{h}} \leq \mathbf{1}^T M \mathbf{1} \leq p\, \lambda_{\max}(M),$$

where $\lambda_{\max}(M)$ is the maximal eigenvalue of M. Notice that the first inequality in the previous expression is saturated in the case where the vector $\mathbf{1}$ is a minimum of the quadratic form $\bar{\mathbf{h}}^T M \bar{\mathbf{h}}$, while the second inequality is saturated when this vector is an eigenvector associated with the maximal eigenvalue.

Since the eigenvalues of M are the inverse-squared singular values of S, the maximal margin of the perceptron is bounded from below by

$$\kappa^* \geq \frac{\sigma_{\min}}{\sqrt{p}}, \tag{4.3}$$

where $\sigma_{\min}$ is the minimal singular value of S.

The simple derivation of equation 4.3, which appears as equation 12 in Tarkowski and Lewenstein (1992), is the main result of our note. This derivation was made possible by the formulation $P2$, which factorizes out the matrix U and allows a one-to-one mapping between the weights $\tilde{\mathbf{w}}$ and the stabilities $\tilde{\mathbf{h}}$. The bound is useful in cases where exact results for the maximal margin are not known.

We also note that for a general S, we cannot improve the bound by providing a simple example where the bound is tight. Consider the following two patterns in a two-dimensional space:

$$\xi^1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \xi^2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \zeta^T = \begin{pmatrix} 1 & -1 \end{pmatrix}. \tag{4.4}$$

In this case the relevant matrices are

$$S = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} -1/\sqrt{2} & -1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{2} \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{4.5}$$

The optimal separating hyperplane for these patterns is the $\xi_2 = 0$ line, and thus the maximal margin is 1. Since $\sigma_{\min} = \sqrt{2}$, the bound 4.3 is tight.

5 A Closed-Form Solution for the Weight Vector

Using the candidate vector $\tilde{\mathbf{h}} = \mathbf{1}$ not only provides us with a bound on the perceptron margin, but can also provide a closed-form solution for the N-dimensional weight vector of $P1$. Using the relationships $\tilde{\mathbf{w}} = \Sigma^{-1} V^T \tilde{\mathbf{h}}$ and $\mathbf{w} = U\tilde{\mathbf{w}}$ in the $r = p$ case defines a weight vector,

$$w_i = z \sum_{\mu, j} \frac{U_{i\mu}}{\sigma_\mu} V_{j\mu}, \tag{5.1}$$

where z is a normalizing factor enforcing $\sum w_i^2 = 1$. The weight vector $\mathbf{w}$ given by equation 5.1 can also be derived by solving for $\mathbf{w}$ in equation 2.3 by applying the pseudo-inverse of S to the vector $\mathbf{h} = \kappa\mathbf{1}$. The result $\mathbf{w} = \kappa S \left(S^T S\right)^{-1} \mathbf{1}$ simplifies using equation 2.4 to equation 5.1 with $z = \kappa$.

This pseudo-inverse solution (Personnaz et al., 1985; Kanter & Sompolinsky, 1987) is not the optimal solution to $P1$, but converges to it as $\kappa \to \infty$. This is because for the optimal $\mathbf{w}$, the fraction of patterns having a margin strictly larger than κ is $\int_{-\infty}^{-\kappa} Dz$, where Dz denotes a gaussian integral (Abbott & Kepler, 1989b), meaning that as κ tends to infinity, the pseudo-inverse solution, equation 5.1, will tend to the optimal solution $\mathbf{w}^*$. Notice that because the error function $\int_{-\infty}^{-\kappa} Dz$ goes exponentially to zero, the pseudo-inverse solution starts approaching the optimal solution for margins κ of order 1. For instance, for $\kappa = 1.40$, less than 5% of the elements of $\mathbf{h}$ are above κ, meaning that $\mathbf{h} \approx \kappa \mathbf{1}$ is already a good approximation.

6 Discussion

We have demonstrated the utility of applying singular value decomposition to the perceptron problem for a quick and simple derivation of several results.

The original problem has N unknowns. However, the patterns actually lie in an r-dimensional subspace spanned by the columns of U, and thus there are only r independent degrees of freedom. The formulation $P2$ uncovers the true dimensionality of the problem by absorbing U in the weight vectors. Another way to look at this result is by noting that the best weight vector $\mathbf{w}$ should be a linear combination of the input patterns (Gerl & Krey, 1994). Indeed, the transformation $\mathbf{w} = U\tilde{\mathbf{w}}$ defines a one-to-one relation between the r-dimensional vectors $\tilde{\mathbf{w}}$ and the N-dimensional vectors $\mathbf{w}$ in the r-dimensional subspace spanned by the patterns. Adding a component to $\mathbf{w}$ that is orthogonal to all patterns will increase the norm of $\mathbf{w}$ without contributing to the classification of the input patterns.

We can characterize the dependence of the weight vector on the input patterns by implicitly defining the vector of pattern contributions $\mathbf{x}$ as $\tilde{\mathbf{w}} = S\mathbf{x}$. The components x_μ are known as the embedding strengths of the patterns and were used to relate the perceptron problem to nonlinear optimization (Anlauf & Biehl, 1989). In particular, for the optimal solution to the perceptron, we have that a given pattern is either exactly on the margin and explicitly encoded by the weights, or it is further away from the margin and is automatically classified without being encoded in the weights. We can easily derive these conditions within our formalism by relating $\mathbf{x}$ and $\bar{\mathbf{h}}$. Specifically, inserting the definitions of $\mathbf{x}$ and M into equation 2.3 implies $\mathbf{x} = M\bar{\mathbf{h}}$. We now note that if we perturb the optimal solution $\bar{\mathbf{h}}$ to equation 4.2 by a vector $\delta\mathbf{h}$, the result is $(\bar{\mathbf{h}} + \delta\mathbf{h})^T M (\bar{\mathbf{h}} + \delta\mathbf{h}) = \bar{\mathbf{h}}^T M\bar{\mathbf{h}} + 2\delta\mathbf{h}^T\mathbf{x} + O(\delta\mathbf{h}^2)$. To ensure the optimality of $\bar{\mathbf{h}}$ in the domain $\{\bar{\mathbf{h}} : \bar{\mathbf{h}} \geq 1\}$, for each μ there are two options: either $\bar{h}_\mu = 1$ and then $\delta h_\mu > 0$, which forces $x_\mu > 0$, or $\bar{h}_\mu > 1$ and then δh_μ can be either positive or negative, which forces $x_\mu = 0$. These conditions are known as the Kuhn-Tucker conditions (Fletcher, 1988; Gerl & Krey, 1994).

Our main result is a simple derivation of a lower bound on the stability margin. This bound becomes tighter as the margin κ increases (Abbott & Kepler, 1989b) and is therefore useful in situations where a large margin is desirable, for instance, in cases where we are interested in increasing the size of the basin of attraction of the fixed points of autoassociative neural networks (Krauth & Mézard, 1987; Forrest, 1988; Gardner & Derrida, 1988; Kepler & Abbott, 1988).

Our analysis up to and including equation 4.1 did not depend on the assumption $r = p$ and is valid also for the cases where $p > N$ (which implies $r < p$). The derivation of equations 4.3 and 5.1, however, does depend on this assumption. In general, solutions to the perceptron may also exist for the case $r < p$. Specifically, in the uncorrelated input case, there exists a solution for $N < 2p$ even though the rank is full only for $N \leq p$. Our methods, however, cannot provide any general results on this regime since there is no one-to-one correspondence between $\bar{\mathbf{h}}$ and $\bar{\mathbf{w}}$. Analysis of this regime has to rely on tools from statistical mechanics such as the replica and cavity methods (Gardner, 1988; Gerl & Krey, 1994; Tarkowski et al., 1991).

Acknowledgments

It is our pleasure to thank Stefano Fusi for helpful discussions and comments on the manuscript. We also thank Misha Tsodyks, Ken Miller, Larry Abbott, Srdjan Ostojic, Michael Vidne, and Dan Rubin for useful suggestions on the manuscript. This research was supported by DARPA grant SyNAPSE HR0011-09-C-0002, the Swartz Foundation, and the Gatsby Foundation. O.B. is supported by the Rothschild Fellowship and the Brainpower for Israel foundation. M.R. is supported by Swiss National Science Foundation grant PBSKP3-133357.

References

Abbott, L., & Kepler, T. (1989a). Optimal learning in neural network memories. *Journal of Physics A: Mathematical and General, 22*, 2031–2038.

Abbott, L., & Kepler, T. (1989b). Universality in the space of interactions for network models. *Journal of Physics A: Mathematical and General, 22*, L711–L717.

Anlauf, J. K., & Biehl, M. (1989). The adatron: An adaptive perceptron algorithm. *EPL (Europhysics Letters), 10*, 687–692.

Brader, J. M., Senn, W., & Fusi, S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Comput., 19*(11), 2881–2912.

Brunel, N., Hakim, V., Isope, P., Nadal, J.-P., & Barbour, B. (2004). Optimal information storage and the distribution of synaptic weights: Perceptron versus Purkinje cell. *Neuron, 43*(5), 745–757.

Fletcher, R. (1988). *Practical methods of optimization*. Hoboken, NJ: Wiley.

Forrest, B. M. (1988). Content-addressability and learning in neural networks. *J. Phys. A: Math. Gen., 21*, 245–255.

Freund, Y., & Schapire, R. (1999). Large margin classification using the perceptron algorithm. *Machine Learning, 37*(3), 277–296.

Gardner, E. (1988). The space of interactions in neural network models. *J. Phys. A, 21*, 257–270.

Gardner, E., & Derrida, B. (1988). Optimal storage properties of neural network models. *J. Phys. A: Math. Gen., 21*, 271–284.

Gerl, F., & Krey, U. (1994). Storage capacity and optimal learning of Potts-model perceptrons by a cavity method. *Journal of Physics A: Mathematical and General, 27*, 7353–7372.

Kanter, I., & Sompolinsky, H. (1987). Associative recall of memory without errors. *Phys. Rev. A, 35*(1), 380–392.

Kepler, T., & Abbott, L. (1988). Domains of attraction in neural networks. *Journal de Physique, 49*(10), 1657–1662.

Korzen, M., & Klesk, P. (2008). Maximal margin estimation with perceptron-like algorithm. In *Artificial Intelligence and Soft Computing ICAISC 2008* (pp. 597–608). Berlin: Springer.

Krauth, W., & Mézard, M. (1987). Learning algorithms with optimal stability in neural networks. *Journal of Physics A: Mathematical and General, 20*(11), L745–L752.

Mézard, M., Parisi, G., & Virasoro, M. (1987). *Spin glass theory and beyond*. Singapore: World Scientific.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Novikoff, A. (1962). On convergence proofs on perceptrons. In *Proceedings of the Symposium on the Mathematical Theory of Automata* (pp. 615–622). Hoboken, NJ: Wiley.

Personnaz, L., Guyon, I., & Dreyfus, G. (1985). Information storage and retrieval in spin-glass like neural networks. *J. Physique Lett., 46*(8), 359–365.

Rigotti, M., Ben Dayan Rubin, D. D., Wang, X.-J., & Fusi, S. (2010). Internal representation of task rules by recurrent dynamics: The importance of the diversity of neural responses. *Frontiers in Computational Neuroscience, 4*(24), 29.

Rosenblatt, F. (1962). *Principles of neurodynamics*. New York: Spartan Books.

Tarkowski, W., Komarnicki, M., & Lewenstein, M. (1991). Optimal storage of invariant sets of patterns in neural network memories. *Journal of Physics A: Mathematical and General, 24*, 4197–4217.

Tarkowski, W., & Lewenstein, M. (1992). Estimates of optimal storage conditions in neural network memories based on random matrix theory. *Journal of Physics A: Mathematical and General, 25*, 6251–6264.

Received November 24, 2010; accepted January 8, 2011.

This article has been cited by: