Covariance Learning of Correlated Patterns in Competitive Networks

Ali A. Minai
Department of Electrical & Computer Engineering and Computer Science
University of Cincinnati
Cincinnati, OH 45221

(To appear in *Neural Computation*)

Covariance learning is a specially powerful type of Hebbian learning, allowing both potentiation and depression of synaptic strength. It is used for associative memory in feed-forward and recurrent neural network paradigms. This letter describes a variant of covariance learning which works particularly well for correlated stimuli in feed-forward networks with competitive K-of-N firing. The rule, which is nonlinear, has an intuitive mathematical interpretation, and simulations presented in this letter demonstrate its utility.

1 Introduction

Covariance-based synaptic modification for associative learning (Sejnowski, 1977) is motivated mainly by the need for synaptic depression as well as potentiation. Long-term depression (LTD) has been reported in biological systems (Levy and Steward, 1983; Stanton and Sejnowski, 1989; Fujii et al., 1991; Dudek and Bear, 1992; Mulkey and Malenka, 1992; Malenka, 1994; Selig et al, 1995), and is computationally necessary to avoid saturation of synaptic strength in associative neural networks.

In the recurrent neural networks literature, covariance learning has been proposed for the storage of biased patterns, i.e., binary patterns with probability(bit $i = 1$) $\neq$ probability(bit $i = 0$) (Tsodyks and Feigel'man, 1991), and this rule is provably optimal among linear rules for the storage of stimulus-response associations between randomly generated patterns under rather general conditions (Dayan and Willshaw, 1991). However, the prescription is not as successful when patterns are correlated, i.e., probability(bit $i = 1$) $\neq$ probability(bit $j = 1$). The correlation causes greater similarity between patterns, making correct association or discrimination more difficult. We present a covariance-type learning rule which works much better in this situation.

The motivation for looking at correlated patterns arises from their ubiquity in neural network settings. For example, if the patterns for an associative memory are characters of the English alphabet on an $n \times n$ grid, some grid locations (e.g., those representing a vertical bar on the left edge) will be much more active than others (e.g., those corresponding to a vertical bar on the right edge). The same argument applies to more complicated stimuli such as images of human faces. The current paper mainly addresses another common source of correlation: processing of (possibly initially uncorrelated) patterns by one or more prior layers of neurons with fixed weights. This situation arises, for example, whenever the output of a "hidden" layer needs to be associated with a set of responses. The results also apply qualitatively to other correlated situations.

The results presented here are limited to the case where all neural firing is competitive K-of-N. This type of firing is widely used in neural modeling to approximate "normalization" of activity in some brain regions

by inhibitory interneurons (e.g., Sharp, 1991; O'Reilly and McClelland, 1994; Burgess et al., 1994). The relationship with the non-competitive case (e.g., Dayan and Wilshaw, 1991) is discussed, but is not the focus of this letter.

2 Methods

The learning problem considered in this letter was to associate m stimulus-response pattern pairs in a two-layer feed-forward network. The stimulus layer was denoted S and the response layer R. All stimulus patterns had N_S bits with K_S 1's and $N_S - K_S$ 0's. Response patterns had N_R bits with K_R 1's and $N_R - K_R$ 0's. This K-of-N firing scheme approximates a competitive situation where broadly directed afferent inhibition keeps overall activity stable. Such competition is likely to occur in the dentate gyrus and CA regions of the mammalian hippocampus (Miles, 1990; Minai and Levy, 1993), which has led to the use of K-of-N firing in several hippocampal models (Sharp, 1991; O'Reilly and McClelland, 1994; Burgess et al., 1994).

The outputs of stimulus neuron j and response neuron i are denoted x_j and z_i, respectively, while w_{ij} denotes the modifiable weight from j to i. The activation of response neuron i is given by:

$$y_i = \sum_{j \in S} w_{ij} x_j \tag{1}$$

All learning was done *off-line*, i.e., the weights were calculated explicitly over a given stimulus-response data set. Performance was tested by presenting each stimulus pattern to a trained network, producing an output through K_R-of-N_R firing, and comparing it with the correct response. A figure of merit for each network on a given data set was calculated as:

$$P = \frac{h - q}{K_R - q} \tag{2}$$

where h is the number of correctly fired neurons averaged over the whole data set and $q \equiv {K_R}^2/N_R$ is the expected number of correct firings for a random response with K_R active bits. Thus, $P = 1$ indicated perfect recall while $P = 0$ represented performance equivalent to random guessing.

Correlated stimulus patterns were generated as follows. First, a $N_S \times N_P$ weight matrix V was generated, such that each entry in it was a 0-mean, unit variance uniform random variable. Then m independent K_P-of-N_P pre-stimulus patterns were generated, and each of these was then linearly transformed by V and the result thresholded by a K_S-of-N_S selection process to produce m stimulus patterns. The correlation arose because of the fixed V weights, and could be controlled by changing the $f_P \equiv K_P/N_P$ fraction — termed the *correlation parameter* in this paper. Figure 1 shows histograms of average activity for individual stimulus neurons in cases with different degrees of correlation. Correlated response patterns were generated similarly with another set of V weights.

3 Mathematical Motivation

The learning rule proposed in this paper is motivated directly by three existing rules: The Hebb rule, the presynaptic rule, and the covariance rule. This motivation is developed systematically in this section. Several other related rules which are also part of this study are briefly discussed at the end of the section.

For binary patterns, insight into various learning rules can be obtained through a probabilistic interpretation of weights. Consider the *Normalized Hebb Rule* defined by:

$$w_{ij} = \langle x_j z_i \rangle \tag{3}$$

where $\langle . \rangle$ denotes an average over all m patterns. The weights in this case can be written as

$$w_{ij} = P(x_j = 1, z_i = 1) \tag{4}$$

i.e., the joint probability of pre- and postsynaptic firing. For correlated patterns, this is not a particularly good prescription.

A significantly better learning rules is the *presynaptic rule*:

$$\begin{aligned} w_{ij} &= \frac{\langle x_j\, z_i \rangle}{\langle x_j \rangle} \\ &= P(z_i = 1 \mid x_j = 1) \end{aligned} \tag{5}$$

This rule is the off-line equivalent of the on-line incremental prescription:

$$w_{ij}(t+1) = w_{ij}(t) + \epsilon x_j(t)(z_i(t) - w_{ij}(t)), \tag{6}$$

hence the designation "presynaptic". This incremental rule has been investigated by several researchers (Minsky and Papert, 1969; Grossberg, 1976; Minai and Levy, 1993; Willshaw et al., 1996). With this rule, the dendritic sum, y_i^μ, of output neuron i when the μth stimulus is presented to the network, is given by:

$$y_i^\mu = \sum_j P(z_i = 1 \mid x_j = 1) x_j^\mu \tag{7}$$

which sums the conditional firing probabilities for i over all input bits active in the μth stimulus. This rule works well in non-competitive situations where each output neuron sets its own firing threshold, and in competitive situations where all output units have identical firing rates. However, in a competitive situation with correlated patterns, output neurons with high firing rates acquire large weights and lock out other output units, precluding proper recall of many patterns.

A rule which addresses the problem of differences in output unit firing rates is the *covariance rule*

$$\begin{aligned} w_{ij} &= \langle x_j\, z_i \rangle - \langle x_j \rangle \langle z_i \rangle \\ &= P(x_j = 1, z_i = 1) - P(x_j = 1) P(z_i = 1) \\ &= P(x_j = 1)[P(z_i = 1 \mid x_j = 1) - P(z_i = 1)] \end{aligned} \tag{8}$$

which accumulates the conditional *increment* in probability of firing over active stimulus bits:

$$y_i^\mu = \sum_j P(x_j = 1)[P(z_i = 1 \mid x_j = 1) - P(z_i = 1)] x_j^\mu \tag{9}$$

However, it also scales the incremental probabilities by $P(x_j = 1)$, which means that highly active stimulus bits contribute more to the dendritic sum than less active ones even if both bits imply the same increment in firing probability. This is not a problem when all stimulus units have the same firing rate, but significantly reduces performance for correlated stimuli.

The above analysis of the presynaptic and covariance rules immediately suggests a hybrid which combines their strengths and reduces their weaknesses. This is the proposed *presynaptic covariance rule*:

$$
\begin{aligned}
w_{ij} &= \frac{\langle x_j z_i \rangle - \langle x_j \rangle \langle z_i \rangle}{\langle x_j \rangle} \\
&= \frac{\langle x_j z_i \rangle}{\langle x_j \rangle} - \langle z_i \rangle \\
&= P(z_i = 1 \mid x_j = 1) - P(z_i = 1) \qquad (10)
\end{aligned}
$$

Now the dendritic sum simply accumulates incremental firing probabilities over all active inputs,

$$
y_i^\mu = \sum_j [P(z_i = 1 \mid x_j = 1) - P(z_i = 1)] x_j^\mu \qquad (11)
$$

and the K_R output units with the highest accumulated increments fire. It can be shown easily that in the K-of-N situation, the mean dendritic sums for the presynaptic, covariance, and presynaptic covariance rules are:

$$
\begin{aligned}
Presynaptic : \langle y_i \rangle &= K_S \langle z_i \rangle \\
Covariance : \langle y_i \rangle &= \sum_j \langle x_j \rangle Cov_{ij} \\
PresynapticCovariance : \langle y_i \rangle &= 0 \qquad (12)
\end{aligned}
$$

where K_S is the number of active bits in each stimulus and Cov_{ij} is the sample covariance of x_j and z_i. Clearly, the presynaptic covariance rule "centers" the dendritic sum to produce fair competition in a correlated situation. It should be noted, however, that there are limits to how much correlation any rule can tolerate; this issue is addressed in the Discussion section below.

3.1 Other Rules

It is also instructive to compare learning rules other than those mentioned above with the presynaptic covariance rule. These are:

1. *Tsodyks-Feigel'man (TF) Rule*:
$$
w_{ij} = \langle (x_j - p)(z_i - r) \rangle \qquad (13)
$$
where $p = {N_S}^{-1} \sum_{j \in S} \langle x_j \rangle$, and $r = {N_R}^{-1} \sum_{i \in R} \langle z_i \rangle$.

2. *Postsynaptic Covariance Rule*:
$$
w_{ij} = \frac{\langle x_j z_i \rangle - \langle x_j \rangle \langle z_i \rangle}{\langle z_i \rangle} \qquad (14)
$$

3. *Willshaw Rule*:
$$
w_{ij} = H(\langle x_j z_i \rangle) \qquad (15)
$$
where $H(y)$ is the Heaviside function.

4. *Correlation Coefficient Rule*:
$$
w_{ij} = \frac{\langle x_j z_i \rangle - \langle x_j \rangle \langle z_i \rangle}{\sigma_j \sigma_i} \qquad (16)
$$
where σ_j and σ_i are sample variances of x_j and z_i, respectively.

The TF rule differs from the covariance rule in its use of global firing rates p and r rather than separate ones for each neuron. The two rules are identical for uncorrelated patterns. Dayan and Willshaw (1991) have shown that the TF rule is optimal among all linear rules for non-competitive networks and uncorrelated patterns.

The postsynaptic covariance rule was included in the list mainly because of its superficial symmetry with the presynaptic covariance rule. In fact, it is obvious that the postsynaptic covariance rule sets weights as

$$w_{ij} = \frac{\langle x_j z_i \rangle}{\langle z_i \rangle} - \langle x_j \rangle = P(x_j = 1 \mid z_i = 1) - P(x_j = 1) = \langle x_j \mid z_i = 1 \rangle - \langle x_j \rangle \tag{17}$$

Thus, the weights to neuron i come to encode the centroid of all stimulus patterns which fire i, offset by the average vector of all stimuli. Essentially, this rule corresponds to a biased version of Kohonen style learning. It is likely to be useful mainly in clustering applications.

The correlation coefficient rule is included because it represents a properly normalized version of the standard covariance rule. The weights are, by definition, correlation coefficients, so the firing of neuron i is based on how positively it correlates with the given stimulus pattern.

Finally, the inclusion of the Willshaw rule reflects its wide use, elegant simplicity and extremely good performance for sparse patterns.

4 Simulation Results

Three different aspects were evaluated via simulations reported here. Unless otherwise indicated, both stimulus and response layers had 200 neurons. N_P was also fixed at 200. All data points were averaged over ten independent runs, each with a new data set. In each run, all rules were applied to the same data set in order to get a fair comparison.

4.1 Pattern Density

In this set of simulations, the number of active stimulus bits, K_S, and active response bits, K_R, for each pattern were varied systematically and the performance of all rules evaluated in each situation. In each run, the stimuli and the responses were correlated through separate, randomly generated weights as described above. The number of associations was fixed at 200. Both stimulus and response patterns were correlated with $f_P = 0.25$.

The results are shown in Table 1. It is clear that the presynaptic covariance rule performs better than any other rule over the whole range of activities. At low stimulus activity, the presynaptic covariance rule is significantly better than the standard covariance rule over the entire range of response activities. As stimulus activity increases, the performance of the presynaptic covariance falls to that of the standard rule. This is because the effect of division by $\langle x_j \rangle$ is attenuated as all average stimulus activities approach 1, making the weights under the presynaptic covariance rule almost uniformly scaled versions of those produced by the standard rule. This suggests that the best situation for using the presynaptic covariance rule is with sparse stimuli. As expected, results with all other rules were disappointing.

4.2 Effect of Correlation

As described earlier, the correlated stimuli and responses used for simulations were generated from random patterns with activity $f_P = K_P/N_P$ via fixed random weights. The higher the fraction f_P, the more correlated the patterns (see Figure 1). To evaluate the impact of different levels of correlation on learning with the presynaptic, covariance, and the presynaptic covariance rules, three values of f_P — 0.25, 0.5, and

0.75 — were investigated. In all cases, N_P was fixed at 200 and network stimuli and responses were 10/200, i.e., quite sparse. The results are shown in Figure 2.

Figure 2(a) shows the effect of stimulus correlation with uncorrelated response patterns. Clearly, all three rules lose some performance, as would be expected from the fact that stimulus correlation corresponds to a loss of information about the outputs. However, the presynaptic covariance rule is significantly superior to the covariance rule. The presynaptic rule, which should theoretically have the same performance as the presynaptic rule in this case is affected by the residual response correlation introduced by finite sample size.

Figure 2(b) shows the impact of response correlation. As expected from theoretical analysis, the covariance and presynaptic covariance rules have almost identical performance, while the presynaptic rule is significantly worse. An interesting feature, however, is the U-shape of the performance curve for the presynaptic rule, which reflects the rule's inherent propensity towards clustering. This is discussed in greater detail below.

4.3 Capacity

To evaluate the capacity of the various learning rules, increasingly large sets of associations were stored in networks of sizes 100, 200 and 400. In each case, the "size" referred to the number of stimulus or response neurons, which was the same. The correlation parameter, f_P, was fixed at 0.25, and stimulus and response activity at 5% active bits per pattern. The results are shown in Figures 3 Figure 4 for correlated stimuli and responses. Again, the presynaptic covariance rule did better than all others at all sizes and loadings, and for all situations.

Graph 4(d) shows how capacity scales with network size for the presynaptic covariance rule. Each line represents the scaling for a specific performance threshold. Clearly, the scaling is linear in all cases, though it is not clear whether this holds for all levels of correlation. Figure 4 shows how the slope of the size-capacity plot changes with performance threshold. The results for the case with correlated stimulus only are also plotted for comparison. The theoretical analysis of these results will be presented in future papers.

5 Discussion

It is instructive to consider the semantics of the presynaptic covariance rule vis-a-vis those of the standard covariance rule in more detail. The weights produced by the presynaptic covariance rule may be written as:

$$w_{ij} = P(z_i = 1 \mid x_j = 1) - P(z_i = 1) = \langle z_i \mid x_j = 1 \rangle - \langle z_i \rangle \tag{18}$$

Thus, a positive weight means that $x_j = 1$ is associated with *higher than average* firing of response neuron i, and a negative weight indicates the reverse. In this way, effects due to disparate activities of stimulus neurons are mitigated, and the firing of response neurons depends only on the average *incremental* excitatory or inhibitory significance of individual stimulus neurons, not just on covariance. In a sense, "cautious" stimulus neurons — those that fire seldom, but when they do fire, reliably indicate response firing — are preferred over "eager" stimulus neurons, which fire more often but often without concomitant response firing. This is very valuable in a competitive firing situation.

To take a somewhat extreme illustrative example, consider a response neuron i with $\langle z_i \rangle = 0.5$, and two stimulus bits x_1 and x_2 with $\langle x_1 \rangle = 0.1$, $\langle x_1 z_i \rangle = 0.1$, and $\langle x_2 \rangle = 0.99$, $\langle x_2 z_i \rangle = 0.5$. Thus, $x_1 = 1 \Rightarrow z_i = 1$ and $z_i = 1 \Rightarrow x_2 = 1$ (or, alternatively, $x_2 = 0 \Rightarrow z_i = 0$). However, $x_2 = 1$ does *not* imply $z_i = 1$. Under the covariance rule, $w_{i1} = w_{i2} = 0.05$, even though $x_1 = 1$ is a reliable indicator of $z_i = 1$ while $x_2 = 1$ is not. The presynaptic covariance rule sets $w_{i1} = 0.5$ and $w_{i2} = 0.0505$, reflecting the increase in the probability of $z_i = 1$ based on each stimulus bit ($x_1 = 1$ raises the probability of $z_i = 1$ to 1.0, while $x_2 = 1$ raises it only to 0.5505.) In this way, each stimulus bit is weighted by its reliability in predicting firing. The information $x = 1 \Rightarrow z = 1$ is extremely valuable even if $x = 0$ does not imply $z = 0$, while $x = 0 \Rightarrow z = 0$ without

$x = 1 \Rightarrow z = 1$ is of only marginal value. Simple covariance treats these two situations symmetrically whereas, in fact, they are quite different. The presynaptic covariance rule rectifies this. However, a price is paid in that correct response firing becomes dependent on just a few cautious stimulus bits, creating greater potential sensitivity to noise. It should be noted that the asymmetry of significance between the two situations just described arises because the usual linear summation of activation treats $x = 1$ and $x = 0$ asymmetrically, discarding the "don't fire" information in $x = 0$.

The comparison with the presynaptic rule is also interesting. As shown Figure 2(b), the presynaptic rule has a U-shaped dependence on output correlation while both the covariance rules have a monotonically decreasing one. This reflects the basic difference in the semantics of the two rule groups. In the presence of output correlation and competitive firing, the presynaptic rule — like the Hebb rule — has a tendency to cluster stimulus patterns, i.e., to fire the same output units for several stimuli. These "winner" outputs are those neurons which have a high average activity. This is a major liability if there are other output neurons with moderately lower average activity, since their firing is depressed disproportionately. However, as output correlation grows to a point where extremely disproportionate firing is the *actual* situation, the clustering effect of both the presynaptic and the Hebb rules ceases to be a liability, and they recover performance. In the extreme case where some response neurons fire for all stimuli and the rest for none, the presynaptic and Hebb rules are perfect. In contrast, the covariance rules fail completely in this situation because all covariances are 0! The main point, however, is that for all but the most extreme response correlation, the presynaptic covariance rule is far better than the presynaptic or the Hebb rules. As pointed out earlier, the presynaptic and the presynaptic covariance rules are identical in the non-competitive situation with adaptive firing thresholds.

5.1 Competitive vs. Non-Competitive Firing

The results of this letter apply only to the competitive firing situation. However, it is interesting to briefly consider the commonly encountered non-competitive situation where each response neuron fires independently based on whether its dendritic sum exceeds a threshold. By assuming that thresholds are adaptive, an elegant signal-to-noise ratio (SNR) analysis can be done to evaluate the performance of different learning rules for threshold neurons (Dayan and Willshaw, 1991). This SNR analysis is based on the dendritic sum distributions for the cases where the neuron fires (the "high" distribution) and those where it does not (the "low" distribution). The optimal firing threshold for the neuron can then be found easily.

The key facilitating assumption in SNR analysis is that each response neuron fires independently. However, this is not so in the competitive case. Indeed, for competitive firing, there is no fixed threshold for any response neuron; an effective threshold is determined implicitly for each stimulus pattern depending on the dendritic sums of all other response neurons in the competitive group. Thus, the distribution of a single neuron's dendritic sum over all stimulus patterns says little about the neuron's performance. It is, rather, the distribution of the dendritic sums of all *neurons* over *each* stimulus pattern which is relevant. The "high" distribution for stimulus k covers the dendritic sums of response neurons which should fire in response to stimulus k, and the "low" distribution covers the dendritic sums of those response neurons that should not fire. The effective firing threshold for stimulus pattern k is then determined by the dendritic sum of the K_Rth most excited neuron. In a sense, there is a duality between the competitive and non-competitive situations. In the former, the goal is to ensure that the correct neurons win for each pattern; in the latter, the concern is that the correct patterns fire a given neuron.

Figure 5 illustrates the results of comparing the errors committed by the covariance and presynaptic covariance rules on the same learning problem in the non-competitive case. The network has 400 neurons in both layers and stores 200 associations. Activity was set at 50 (25%) to ensure reasonable sample size, and f_P was set to 0.25. The threshold was set at the intersection point of the empirically determined "high" and "low" dendritic sum distributions. This is optimal for Gaussian distributions. Errors of omission (0 instead of 1) and commission (1 instead of 0) are shown separately. While Figure 5 is based only on a single run, it does suggest that the presynaptic covariance rule improves on the covariance rule in the non-competitive case as well.

6 Conclusion

The results presented above demonstrate that, for correlated patterns in competitive networks, the presynaptic covariance rule is clearly superior to the presynaptic rule or the standard covariance rule when the stimulus patterns are relatively sparsely coded (average activity < 0.5). This is true with respect to capacity as well as performance. However, these conclusions do need to be formally verified through mathematical analysis and more simulations. The results of these will be reported in future papers. For uncorrelated stimulus patterns, simulations (not shown) indicate that the standard covariance rules has virtually the same performance as the presynaptic covariance rule.

References

Burgess, N., Recce, M., and O'Keefe, J. (1994) A Model of Hippocampal Function. *Neural Networks* **7:** 1065-1081.

Dayan, P., and Willshaw, D.J. (1991) Optimising Synaptic Learning Rules in Linear Associative Memories. *Biol. Cybernetics* **65:** 253-265.

Fujii, S., Saito, K., Miyakawa, H., Ito, K-I., and Kato, H. (1991) Reversal of Long-Term Potentiation (Depotentiation) Induced by Tetanus Stimulation of the Input to CA1 Neurons of Guinea Pig Hippocampal Slices. *Brain Res.* **555:** 112-122.

Green, T.S., and Minai, A.A. (1995a) A Covariance-Based Learning Rule for the Hippocampus. *4th Computational Neural Systems Meeting, Monterey, CA.* (to appear).

Green, T.S., and Minai, A.A. (1995b) Place Learning with a Stable Covariance-Based Learning Rule. *Proc. WCNN* **I:** 573-576.

Grossberg, S. (1976) Adaptive Pattern Classification and Universal Recoding: I. Parallel Development an Coding of Neural Feature Detectors. *Biol. Cybernetics* **23:** 121-134.

Hebb, D.O. (1949) *The Organization of Behavior.* New York: Wiley.

Levy, W.B., and Steward, O. (1983) Temporal Contiguity Requirements for Long-Term Associative Potentiation/Depression in the Hippocampus. *Neuroscience* **8:** 791-797.

Malenka, R.C. (1994) Synaptic Plasticity in the Hippocampus: LTP and LTD. *Cell* **78:** 535-538.

Minai, A.A., and Levy, W.B. (1993) Sequence Learning in a Single Trial. *Proc. WCNN***II:** 505-508.

Minai, A.A., and Levy, W.B. (1994) Setting the Activity Level in Sparse Random Networks. *Neural Computation* **6:** 85-99.

Minsky, M., and Papert, S. (1969) **Perceptons**. MIT Press, Cambridge, MA.

Mulkey, R.M., and Malenka, R.C. (1992) Mechanisms Underlying Induction of Homosynaptic Long-Term Depression in Area CA1 of the Hippocampus. *Neuron* **9:** 967-975.

O'Reilly, R.C., and McClelland, J.L. (1994) Hippocampal Conjunctive Encoding, Storage, and Recall: Avoiding a Tradeoff. *Hippocampus* **6:** 661-682.

Sejnowski, T.J. (1977) Storing Covariance with Nonlinearly Interacting Neurons. *J. Math. Biol.* **4:** 303-321.

Selig, D.K., Hjelmstand, G.O., Herron, C., Nicoll, R.A., and Malenka, R.C. (1995) Independent Mechanisms for Long-Term Depression of AMPA and NMDA Responses. *Neuron* **15:** 417-426.

Sharp, P.E. (1991) Computer Simulation of Hippocampal Place Cells. *Psychobiology* **19:** 103-115.

Stanton, P.K., and Sejnowski, T.J. (1989) Associative Long-Term Depression in the Hippocampus Induced by Hebbian Covariance. *Nature* **339:** 215-218.

Tsodyks, M.V., and Feigel'man, M.V. (1988) The Enhanced Storage Capacity in Neural Networks with Low Activity Level. *Europhys. Lett.* **6:** 101-105.

Willshaw, D.J., Buneman, O.P., and Longuet-Higgins, H.C. (1969) Non-Holographic Associative Memory. *Nature* **222:** 960-962.

Willshaw, D., Hallam, J., Gingell, S., and Lau, S.L. (1996) Marr's Theory of the Neocortex as a Self-Organising Neural Network. *Preprint*, University of Edinburgh, UK.

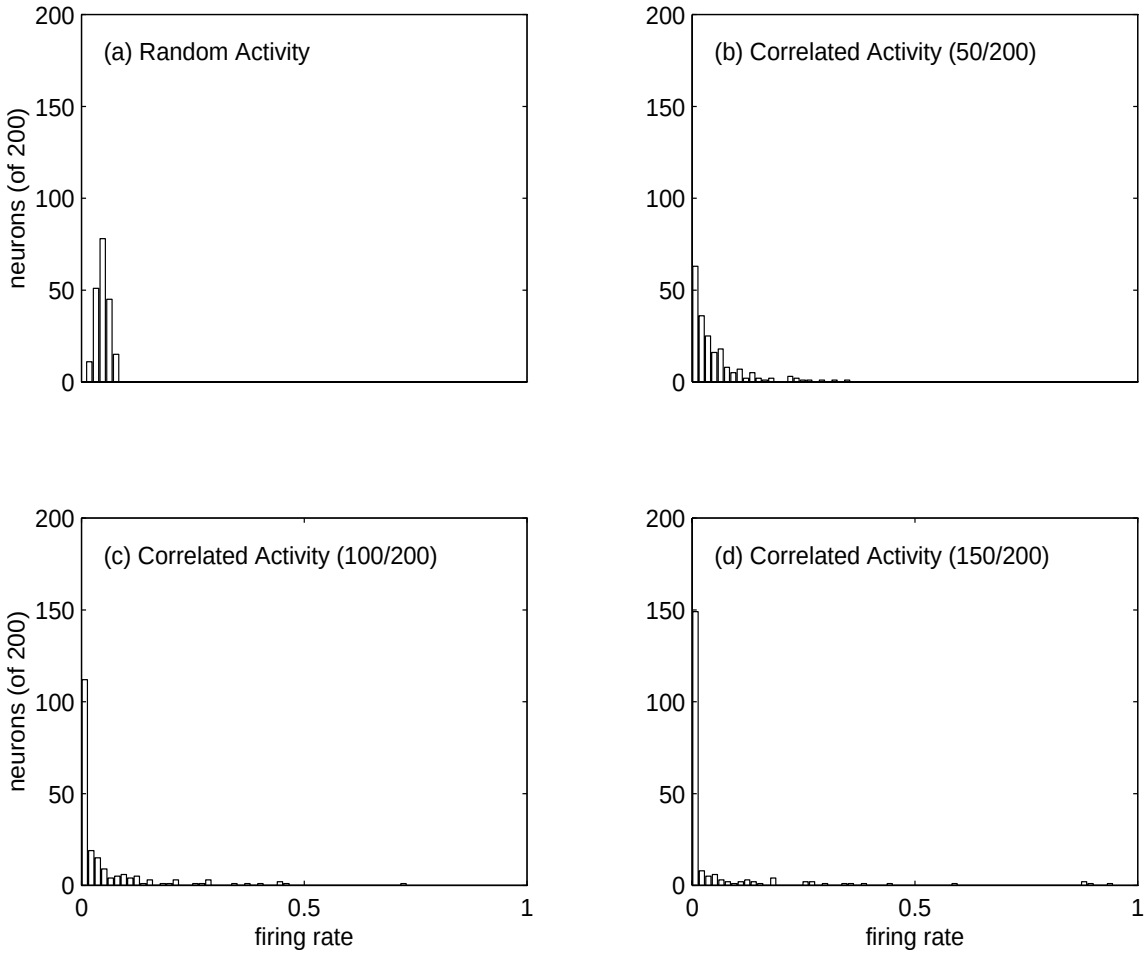

Figure 1: Each graph shows the histogram of individual neuron activities (i.e., fraction of patterns for which a neuron is 1) for four different levels of correlation: (a) Randomly generated patterns, no correlation; (b) 50 of 200 pre-stimulus neurons active per pattern ($f_P = 0.25$); (c) 100 of 200 pre-stimulus neurons active ($f_P = 0.5$); (d) 150 of 200 pre-stimulus neurons active ($f_P = 0.75$). The distribution of activities is close to normal for random patterns, but becomes increasingly skewed with growing correlation.

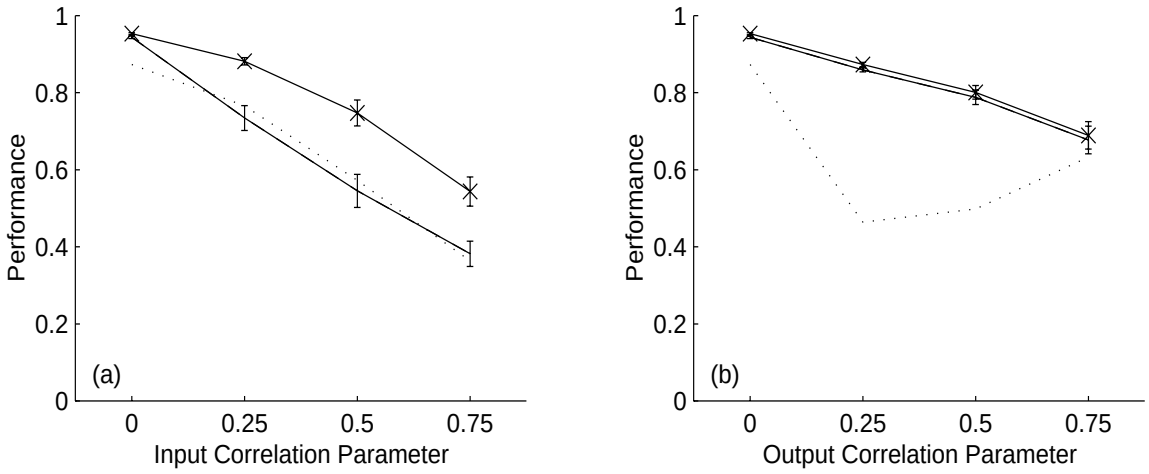

Figure 2: The performance of the learning rules at different correlation levels for stimulus and response correlations. In each graph, the $\times$ symbol indicates the performance of the presynaptic covariance rule, the plain solid line the performance of the standard covariance rule, and the dotted line that of the presynaptic rule. Graph (a) is for the case of correlated stimuli and uncorrelated responses, while Graph (b) shows the results for correlated responses and uncorrelated stimuli. In all cases, stimulus and response activities were set at 10-of-200. N_P was set at 200 and K_P at 50. Each data point was averaged over ten independent runs. The network had $N_S = N_R = 200$, and 200 associations were stored during each run.

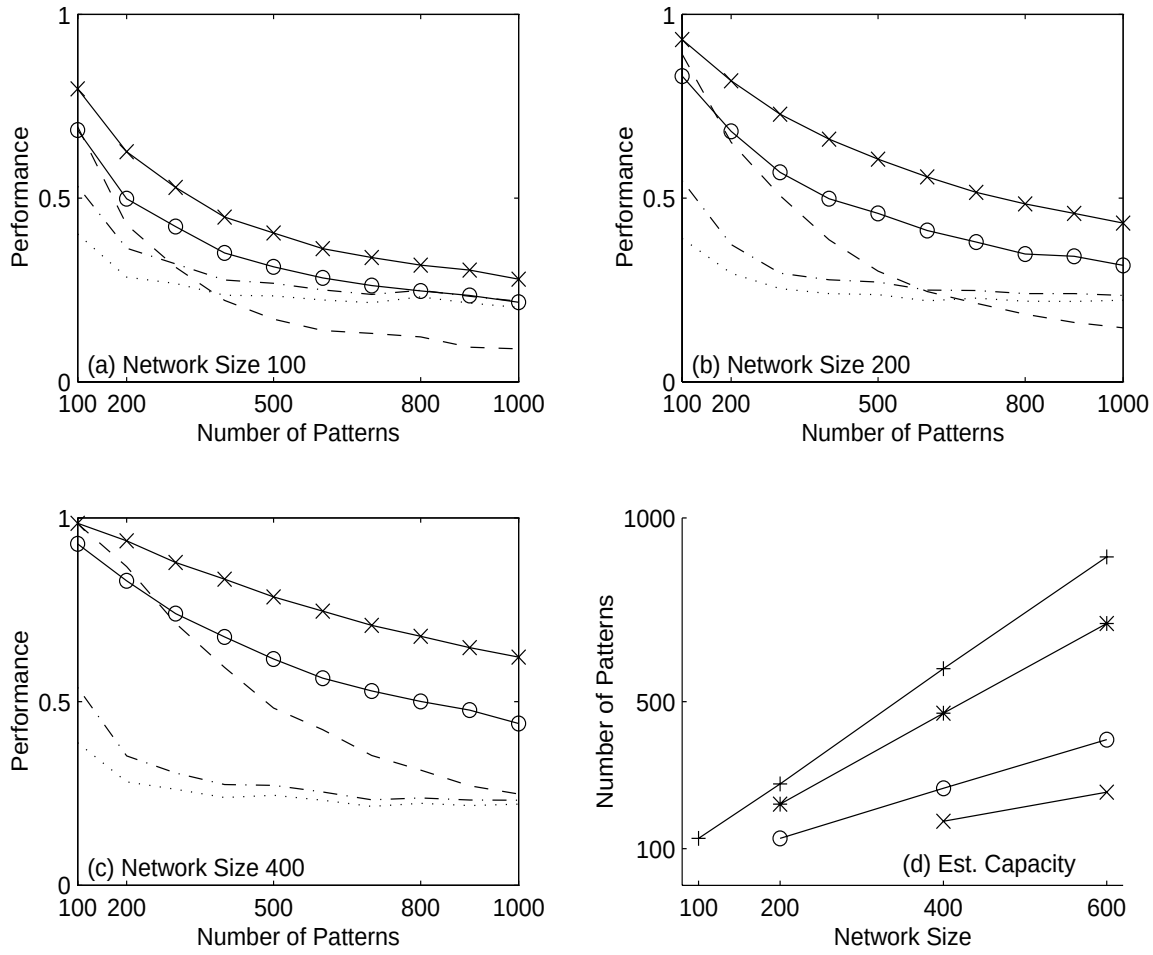

Figure 3: The performance of the learning rules for data sets and networks of different sizes. bias levels. In Graphs (a)-(c), the dashed line shows the performance of the Willshaw rule, the dot-dashed line that of the TF rule, and the dotted line the performance of the normalized Hebb rule. The performance of the presynaptic covariance rule is indicated by $\times$ and that of the standard covariance rule by $\circ$. Graph (d) shows the empirically determined dependence of the presynaptic rule's capacity as a function of network size for performance levels of $0.95(\times)$, $0.9(\circ)$, $0.8(*)$, and $0.75(+)$. In all cases, stimulus and response activities were set at 10-of-200. N_P was set at 200 and K_P at 50 for both stimuli and responses. Each data point was averaged over ten independent runs.

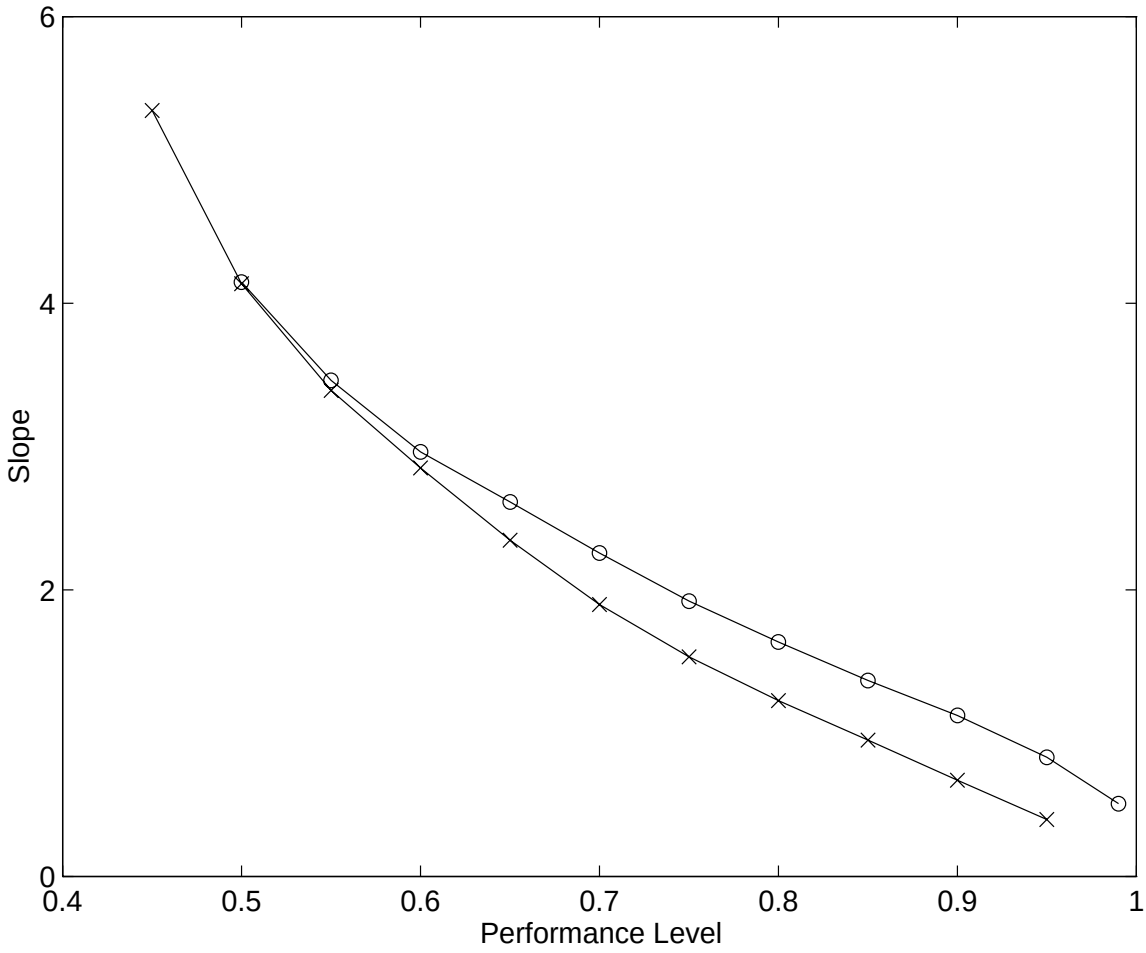

Figure 4: The graph shows the slope of the capcity-size line for the presynaptic covariance rule (cf. Figure 3(d)) as a function of the performance level used to determine capacity. The curve with $\times$ markers corresponds to the case reported in Figure 3 while the curve with $\circ$ markers is for correlated stimuli and uncorrelated responses. The latter is included for comparison purposes. All parameters are as in Figure 3.

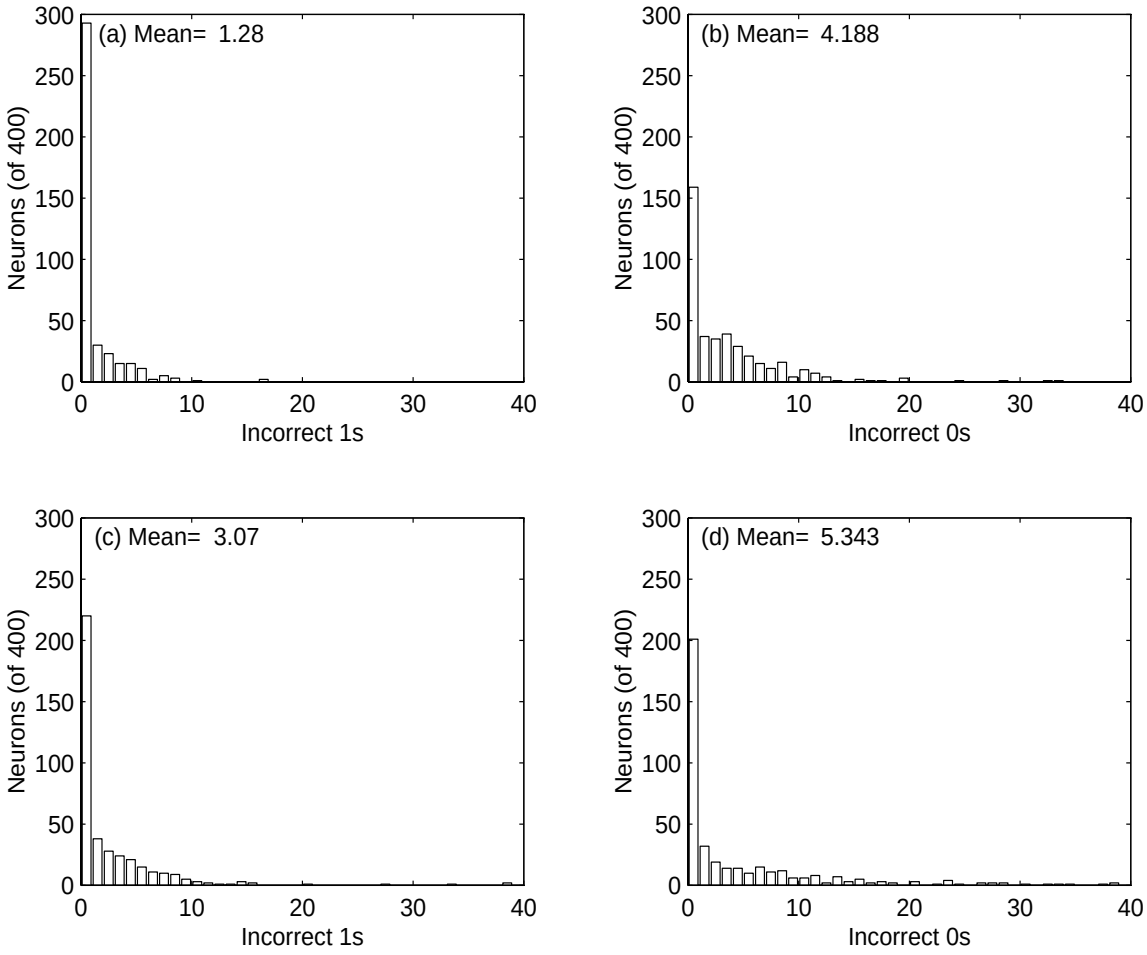

Figure 5: Errors of omission and commission by the presynaptic covariance and the standard covariance rule in a single run with non-competitive firing. The threshold for each neuron was set at the empirically determined intersection point of its "low" and "high" dendritic sum distributions. Graphs (a) and (b) show the histograms for errors of commission (a) and omission (b) for the presynaptic covariance rule, and graphs (c) and (d) show the same for the standrard covariance rule. The networks have $N_S = N_R = 400$, and $K_S = K_R = 50$. The correlation parameter is 0.25 for both stimuli and responses, and 200 associations are stored in each case.

Rule	10/10	10/100	10/190	100/10	100/100	100/190	190/10	190/100	190/190
Presynaptic Covariance	0.8211	0.6012	0.8236	0.8032	0.5202	0.7969	0.6518	0.4182	0.6623
	±0.0129	±0.0131	±0.0068	±0.0056	±0.0091	±0.0092	±0.0363	±0.0180	±0.0272
Covariance	0.6676	0.4337	0.6700	0.7936	0.5117	0.7861	0.6776	0.4358	0.6805
	±0.0295	±0.0207	±0.0192	±0.0086	±0.0133	±0.0115	±0.0302	±0.0187	±0.0232
Presynaptic	0.4213	0.4717	0.4320	0.2116	0.3434	0.2163	0.2082	0.3485	0.2035
	±0.0156	±0.0284	±0.0219	±0.0172	±0.0157	±0.0321	±0.0120	±0.0158	±0.0188
Hebb	0.2854	0.4006	0.2768	0.2051	0.3480	0.2117	0.2087	0.3485	0.2131
	±0.0326	±0.0123	±0.0179	±0.0187	±0.0183	±0.0330	±0.0120	±0.0159	±0.0243
Tsodyks-Feigelman	0.3675	0.4467	0.3439	0.2723	0.3842	0.2719	0.3596	0.4558	0.3673
	±0.0535	±0.0236	±0.0252	±0.0156	±0.0188	±0.0215	±0.0476	±0.0090	±0.0270
Postsynaptic Covariance	0.2157	0.4340	0.6698	0.2860	0.5112	0.7926	0.2198	0.4362	0.6854
	±0.0232	±0.0214	±0.0203	±0.0404	±0.0138	±0.0083	±0.0206	±0.0184	±0.0247
Willshaw	0.6761	0.2856	0.1247	0.1187	0.0082	0.0066	0.0227	-0.0014	0.0074
	±0.0264	±0.0243	±0.0231	±0.0102	±0.0296	±0.0164	±0.0211	±0.0220	±0.0270
Correlation Coefficient	0.6306	0.5276	0.6452	0.6654	0.5075	0.6570	0.3190	0.2567	0.3600
	±0.0263	±0.0163	±0.0182	±0.0165	±0.0136	±0.0338	±0.0378	±0.0412	±0.0387

Table 1: The performance of each rule at different levels of stimulus and response coding densities. The network has $N_S = N_R = 200$, and 200 stored associations in each run. Each rule was evaluated in nine situations: Stimulus activities of 10, 100, and 190 out of 200, each at response activities of 10, 100, and 190 out of 200. These are indicated in the top row of the table by stimulus/response activity. The ± entries indicate sample standard deviation in each case. N_P was set at 200 and K_P at 50 for both stimuli and responses. Each data point is averaged over ten independent runs.