A Comparative Analysis of Hebbian and Widrow-Hoff Learning

Michael Balas, *Student, Faculty of Health Sciences*

Abstract—This paper seeks to build upon existing literature by offering quantitative comparisons in the performance between the Hebbian and Widrow-Hoff (delta) learning algorithms for training Hetero-Associative neural networks to recall binary patterns. The delta rule was found to be 21% more accurate than the Hebb rule on average after 10 patterns, upon which both networks begin producing severe recall errors. The delta rule became more accurate with increasing input neurons, while the opposite effect was observed with the Hebb rule. The number of output neurons did not affect the accuracy of either model. Based on the results presented in this paper, along with similar findings in previous studies, the delta rule is a much more effective learning algorithm.

Index Terms—Hebb rule, Widrow-Hoff, delta rule, pattern associator, neural network.

1 Introduction

The Pattern Associator is one of the simplest types of neural networks, providing a basic model of the human brain by storing and associating similar patterns. These networks consist of a single layer of input and output units, x and y, respectively, as well as a weight matrix, W, representing the connections from x to y[1]. The weights are determined in such a way to store a set of P pattern associations, *W(P)*[1]. There are two types of associative memory networks: **(i)** Auto-Associative Memory, where the input and output vectors are the same, and **(ii)** Hetero-Associative Memory, where they are different. In other words, the vector *x(P)* is composed of n-components (n-tuple), while the vector *y(P)* is composed of m-components (m-tuple). In this paper, feedforward Hetero-Associative Memory neural networks will be explored, with the patterns represented as a binary system of *on* (1) and *off* (0) signals. The architecture of this model is shown in Figure 1.

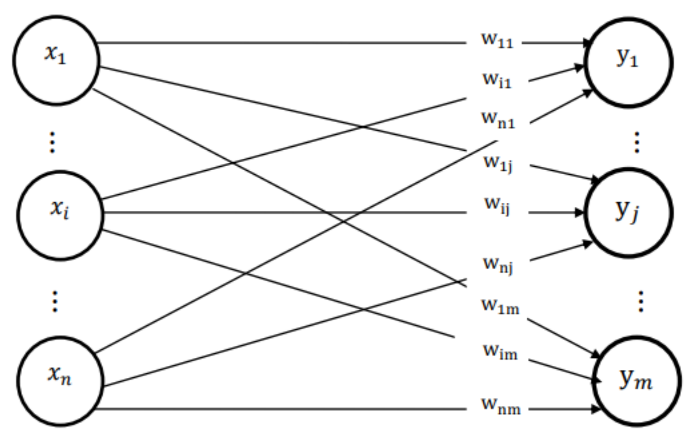

Fig. 1. Architecture of the feedforward Hetero-Associative neural network

The two learning algorithms that will be used for determining the weights and achieving pattern association are the Hebb rule and the Widrow-Hoff rule, more commonly known and henceforth referred to as the delta rule. Hebbian learning, introduced in 1949 by Donald Hebb in *The Organization of Behavior*, is the simplest and first learning algorithm that has been used for unsupervised neural networks[2]. Hebb proposed that the synaptic strength (weight) increases when both source and destination neurons are activated, stated simply as, *cells that fire together wire together*. This postulate can be mathematically expressed as:

$$\Delta w_{ij} = \alpha x_i y_j$$

where $i \in [1..n]$, $j \in [1..m]$ and α refers to the learning rate parameter.

Developed by Widrow and Hoff in 1960, the delta rule is a common supervised learning algorithm that iteratively adjusts the weights of neural connections to minimize the difference between the desired and the obtained signals (in other words, minimizing error signal over all training patterns)[3]. The delta rule is given by:

$$\Delta w_{ij} = \alpha(t_j - y_j)x_i$$

where t_j refers to the target output, α refers to the learning rate parameter, $i \in [1..n]$, $j \in [1..m]$, and $y_j = \sum_{i=1}^{n} x_i w_{ij}$.

2 Methods

Simulations were performed using MATLAB R2017b on an HP EliteBook Folio 9480m computer. Performance of the neural networks are measured using the *Hamming distance* between the output and target neurons[4]. Given two vectors in a binary field, $u, v \in \mathbb{F}_2$, the Hamming distance, $H(u, v)$, is defined as the number of positions where u and v differ, and can be thought of as the number of bits needed to change u to v or vice versa[4]. This value is then divided by the product of the number of output neurons and number of patterns (the number of elements in the output or target neuron matrices), and subtracted from 1 to give the percent

• *M. Balas is studying under the supervision of Dr. Suzanna Becker in the Neurotechnology & Neuroplasticity Lab at McMaster University.*
E-mail: 1michaelbalas@gmail.com

Manuscript received March 08, 2018

accuracy of the output patterns relative to the desired output.

The Hebbian weights are initialized by a zero matrix, whereas the delta weights are initialized by a matrix of randomly generated integers ranging from 0-10. The Hebbian learning model does not use a learning rate (α=1), whereas the delta learning rate is equal to the reciprocal of the number of input neurons and runs for 1000 epochs. Once the weights are obtained, both models use a linear threshold activation function to produce the output patterns:

$$f(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

Three experiments were conducted to assess the difference in output pattern accuracy between Hebbian learning and the delta rule by measuring the effects of **(i)** number of patterns; **(ii)** number of input neurons; and **(iii)** number of output neurons. Each experiment ran 1000 simulations to observe the average effects and all other non-tested parameters were kept constant. The inputs and target patterns were randomly generated binary matrices with sizes varying upon experiment.

3 Results

3.1 The Effects of Increasing Patterns

Both learning models experience significant reductions in output accuracy with increasing numbers of patterns (Fig. 2). Each pattern was investigated using 10 input neurons and two output neurons. The delta rule exhibits perfect or near-perfect accuracy with the first few (2-5) patterns, whereas the Hebb rule is already, on average, less than 80% accurate with only two patterns. Both networks demonstrate significant cross-talk when having to learn many (>5) patterns, until the point they both reach saturation and recall errors become severe. This is evident in the way both networks taper off and remain at 50% accuracy when having to recall many patterns, which is simply the random chance of acquiring the correct outcome in a binary field. Both learning algorithms reach saturation after approximately 10 patterns, though this is less evident for the Hebbian network since it is already within 50-55% accuracy after only five patterns (the delta rule maintains above 90% accuracy with five patterns). Interestingly, this network saturation is marked by an all-ones matrix in the Hebbian model, and by a zero matrix in the delta model.

3.2 The Effects of Increasing Input Neurons

The Hebbian and delta learning algorithms demonstrated opposite effects with increasing numbers of input neurons (the number of output neurons was kept constant at 2). Among the three tested cases (N=4, N=10, and N=16), the delta rule achieved the highest accuracy throughout with 16 input neurons, while 10 inputs were slightly lower and 4 inputs were significantly worse, with 12% lower accuracy per number of patterns on average (before reaching saturation) (Fig. 3). In contrast, the Hebbian model consistently achieved the best results with the least amount of input neurons, while 10 and 16 inputs gave on average 4% lower accuracy, (with 16 being faintly worse than 10) (Fig. 4). For both models, network saturation occurred at a pattern number of 10, regardless of the number of input neurons.

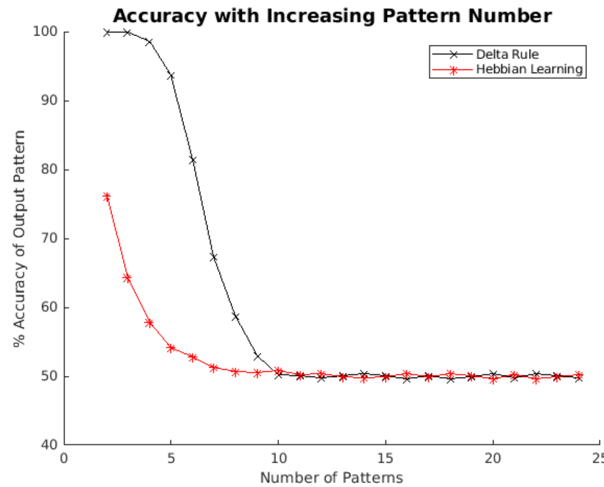

Fig. 2. The delta rule demonstrates significantly better pattern recall until both learning algorithms reach network saturation at about 10 patterns

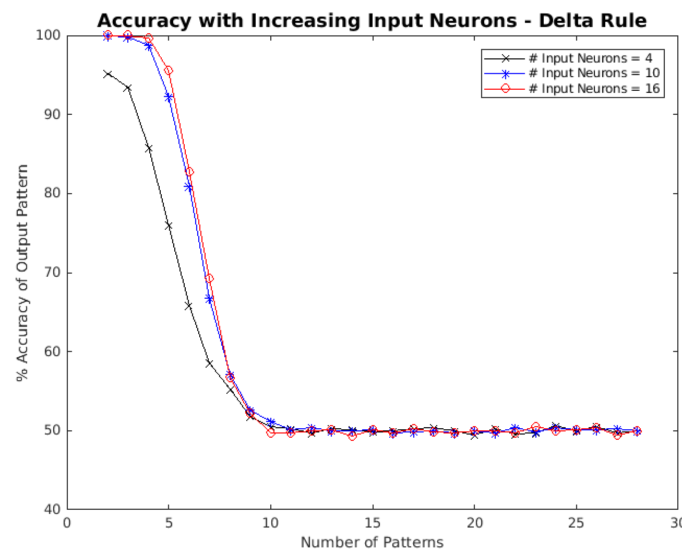

Fig. 3. The delta rule exhibits significantly increased accuracy with greater numbers of input neurons

3.3 The Effects of Increasing Output Neurons

The number of output neurons had no bearing on output pattern accuracy for both Hebbian and delta learning algorithms (tested cases of 2, 6, 10 and 14 output neurons while inputs remained constant at 10) (Fig. 5 & Fig. 6).

4 Discussion

The Hebb rule demonstrates inaccurate pattern recall since it is generally not possible to avoid cross-talk between the responses to different patterns, as this requires the training input to form an orthogonal set. The delta rule achieves better accuracy because it overcomes the orthogonality limitation imposed by the Hebb rule, though it still requires linearly independent inputs for accurate pattern recall.

One possible method for yielding improved accuracy is using bipolar rather than binary representations of training

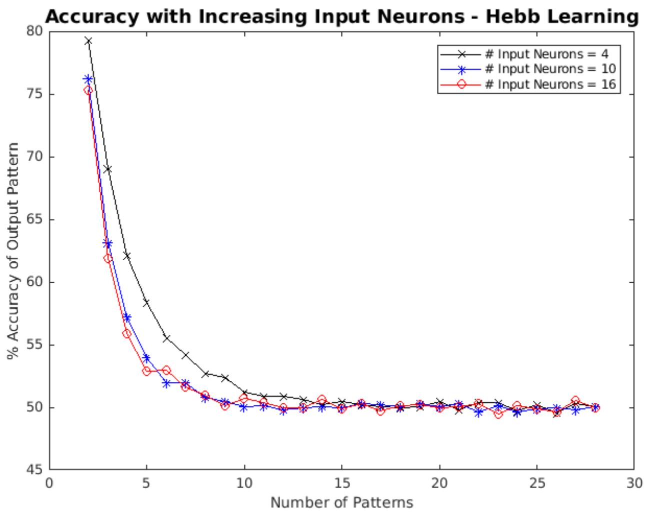

Fig. 4. The Hebb rule displays moderately improved pattern recall with smaller numbers of input neurons

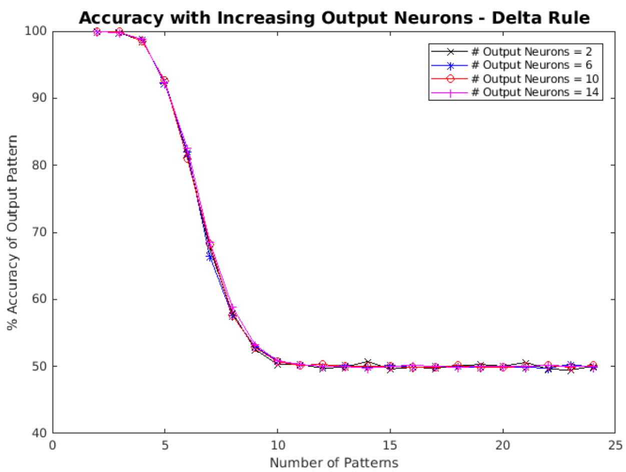

Fig. 5. No significant change in accuracy occurred with varying output neuron numbers in the delta learning models

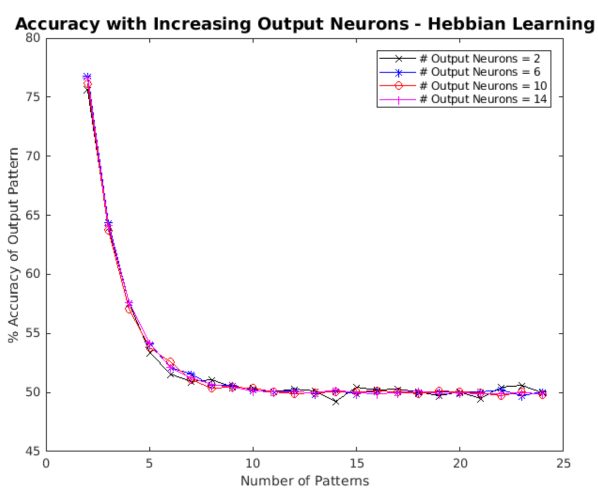

Fig. 6. No significant change in accuracy occurred with varying output neuron numbers in the Hebbian learning models

patterns. Previous studies have already demonstrated that bipolar representations are more robust in the presence of noise, and are better in terms of strength and sign of correction coefficients[5,6]. Moreover, to prevent infinitely large weight matrices which are typical of Hebbian learning models, modified learning algorithms, such as Oja's normalization rule, can be used[7].

In summary, the findings presented in this paper demonstrate that the delta rule is significantly more accurate than Hebbian learning, despite both models reaching network saturation at 10 patterns. Furthermore, the number of input neurons enhanced the accuracy of the delta model, while exhibiting the opposite effects in the Hebbian model. The magnitude of these effects were also three times greater in the delta compared to the Hebbian model. Output neuron count showed no effect on accuracy in either model.

Acknowledgment

The author would like to thank Dr. Suzanna Becker for her guidance and support in the undertaking of this project.

References

[1] Rumelhart DE, McClelland JL, PDP Research Group. Parallel distributed processing. Cambridge, MA: MIT press; 1987 Jul.
[2] Hebb DO. The organization of behavior: A neuropsychological theory. Psychology Press; 2005 Apr 11.
[3] Widrow B, Hoff ME. Adaptive switching circuits. Standford Univ CA Electronics Labs; 1960 Jun 30.
[4] Hosangadi S. Distance measures for sequences. arXiv preprint arXiv:1208.5713. 2012 Aug 28.
[5] Fausett L. Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall, Inc.; 1994 Jul 1.
[6] Kosko B. Bidirectional associative memories. IEEE Transactions on Systems, man, and Cybernetics. 1988 Jan;18(1):49-60.
[7] Hyvrinen A, Oja E. One-unit learning rules for independent component analysis. InAdvances in neural information processing systems 1997 (pp. 480-486).