

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**Redes Neuronales Inspiradas Biológicamente para el
Reconocimiento Dinámico de Patrones**

Miguel Francisco Mena Larrea

Matemáticas

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Matemático

Quito, 7 de mayo de 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

**Redes Neuronales Inspiradas Biológicamente para el Reconocimiento Dinámico de
Patrones**

Miguel Francisco Mena Larrea

Julio César Ibarra Fiallo, Ph.D. (c)

.....

Antonio Di Teodoro, Ph.D.

.....

Quito, 7 de mayo de 2025

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Miguel Francisco Mena Larrea

Código: 00335649

C.I.: 1105233777

Fecha: Quito, 7 de mayo de 2025

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>

RESUMEN

Este trabajo presenta un estudio sobre redes neuronales biológicamente inspiradas para el reconocimiento dinámico de patrones visuales, tomando como referencia los modelos desarrollados por Stephen Grossberg. Se analiza en detalle la arquitectura de la Red de Grossberg, que combina mecanismos de memoria a corto y largo plazo, normalización de señales y realce de contraste, inspirados en el funcionamiento del sistema visual humano. A continuación, se introduce la Teoría de Resonancia Adaptativa (ART) como una solución al problema de estabilidad en el aprendizaje incremental continuo. ART incorpora un mecanismo de expectativas basado en el sistema olfativo. Finalmente, se implementa y evalúa el simplificado modelo ART-1, mostrando su rendimiento en tareas de clasificación sobre el conjunto MNIST, así como estrategias adicionales para mitigar problemas de sobresegmentación y mejorar su aplicabilidad en dominios complejos.

Palabras clave: Redes Neuronales Recurrentes, Modelo de Shunting, Red de Grossberg, Adaptive Resonance Theory, Sistema Visual, Sistema Olfativo Conjuntos, Reconocimiento Dinámico de Patrones.

ABSTRACT

This work presents a study on biologically inspired neural networks for the dynamic recognition of visual patterns, using the models developed by Stephen Grossberg as a reference. It provides a detailed analysis of the Grossberg Network architecture, which integrates short- and long-term memory mechanisms, signal normalization, and contrast enhancement, all inspired by the functioning of the human visual system. Subsequently, the Adaptive Resonance Theory (ART) is introduced as a solution to the stability problem inherent in continuous incremental learning. ART incorporates an expectation mechanism based on the olfactory system. Finally, the simplified ART-1 model is implemented and evaluated, demonstrating its performance on classification tasks using the MNIST dataset, along with additional strategies to mitigate over-segmentation issues and enhance its applicability to complex domains.

Keywords: Recurrent Neural Networks, Shunting Model, Grossberg Network, Adaptive Resonance Theory, Visual System, Olfactory System, Dynamic Pattern Recognition.

ÍNDICE GENERAL

1	Introducción	11
1.1	Motivación Biológica	12
1.2	Redes Neuronales Recurrentes	14
1.3	El Modelo de <i>Shunting</i>	15
2	Red de Grossberg	17
2.1	Layer 1: Retina	18
2.2	Layer 2: Corteza Visual y STM	19
2.3	Pesos adaptativos: LTM	20
3	Adaptive Resonance Theory	22
3.1	Layer 1: Bulbo Olfativo	23
3.2	Layer 2: Corteza Olfativa y STM	24
3.3	Susbsistema de Orientación	24
3.4	Pesos adaptativos: Expectativa y LTM	25
4	Experimentos	27
4.1	Algoritmo ART-1	27
4.2	Resultados en Datos Artificiales	28
4.3	Resultados en MNIST	29
4.3.1	Testores Típicos	31
4.3.2	El Problema del Sesgo Inicial	32
4.3.3	Estrategias de Inicialización Supervisada	33
4.3.4	Comparación con Métodos de Referencia	35
5	Conclusiones	36
Bibliografía		38

ÍNDICE DE FIGURAS

1.1	Puntos ciegos en la retina. Imagen tomada de [1].	12
1.2	Triángulo y círculo ilusorios debido a los procesos del sistema visual. Imagen tomada de [1].	13
1.3	Ilusión del vestido azul y negro o blanco y dorado. Imagen tomada de [2].	13
1.4	<i>Leaky integrator</i> $n(0) = 1$, estímulo constante $p = 2$ y distintas velocidades de decaimiento ε	15
2.1	Arquitectura de Red de Grossberg. Imagen tomada de [1].	17
2.2	Competencia de dos neuronas en L2. Imagen adaptada de [1].	20
3.1	Arquitectura de Red ART. Imagen tomada de [1].	22
4.1	Letras A, B, C y D clasificadas con parámetros de vigilancia $\rho = 0.9, 0.7$ y 0.4	28
4.2	Distribución del número de clases generadas por ART-1 para varios ρ en el subconjunto MNIST de 350 imágenes.	30
4.3	Distribución del número de clases generadas por ART-1 para varios ρ en el subconjunto MNIST de 350 imágenes.	31
4.4	Primeros 10 inputs de MNIST sobre los que entrena el modelo ART-1	32
4.5	Patrones $\mathbf{W}^{2:1}$ almacenados en la LTM del modelo ART-1 después de entrenar todo MNIST	32
4.6	Pesos iniciales $\mathbf{W}^{1:2}$, $\mathbf{W}^{2:1}$ más cercanos al centroide de cada clase.	34

ÍNDICE DE CUADROS

4.1	Precisión alcanzada por el modelo ART-1 bajo diferentes estrategias de inicialización.	35
4.2	Comparación de precisión reportada en MNIST por modelos supervisados y estáticos.	35

DEDICATORIA

Para mis padres, para mis hermanas, para Belén y para quien le sea útil.

AGRADECIMIENTO

Quiero expresar mi profundo agradecimiento a mi tutor de tesis, Julio Ibarra, por su invaluable orientación, apoyo incondicional y paciencia durante todo el proceso de investigación.

Además quiero agradecer todas las personas que han sido parte de mi vida, mis padres, mis hermanas, mis abuelos, mi familia, mis amigos y Belén.

CAPÍTULO 1

INTRODUCCIÓN

La Red de Grossberg es un tipo de red neuronal recurrente inspirada en procesos cognitivos humanos introducida por Stephen Grossberg en 1976 [3]. Su trabajo fue pionero en la integración de neurociencia y aprendizaje automatizado, estableciendo un puente entre las ciencias cognitivas y los modelos computacionales. Grossberg sentó las bases teóricas para el desarrollo de modelos que imitan el funcionamiento del cerebro humano, mucho antes de la aparición del *deep learning* actual.

El objetivo de la Red de Grossberg es replicar las propiedades experimentales observadas en el sistema visual humano, especialmente su extraordinaria capacidad de adaptación y reconocimiento frente a estímulos variados y dinámicos. Su primer modelo formal [4] describe una red bicapa recurrente con capacidad de “recodificación universal”, es decir, capaz de recodificar un conjunto arbitrariamente grande y variado de patrones. Este logro constituye un avance teórico significativo, al demostrar que la red posee una plasticidad lo suficientemente amplia como para manejar patrones altamente diversos y llevar a cabo recodificaciones no triviales, posicionándose como un modelo de referencia en el estudio de redes neuronales que imitan el funcionamiento del cerebro humano.

Sin embargo, el modelo es inestable. Si a lo largo del tiempo un número arbitrario de patrones perturban un conjunto fijo de neuronas, los clústers formados nunca logran estabilizarse. En otras palabras, el sistema no puede incorporar nueva información sin comprometer lo previamente aprendido. De esta problemática surge ART (del inglés *Adaptive Resonance Theory*), un segundo modelo [5] inspirado en el sistema olfativo humano, que utiliza un mecanismo de “resonancia” para moderar y estabilizar la clasificación frente a un flujo constante de información.

1.1. Motivación Biológica

La base de la Red de Grossberg está fuertemente inspirada en el sistema visual humano, compuesto por córnea, retina, nervio óptico y corteza visual, encargados de procesar la luz que ingresa al ojo para transformarla en una imagen que el cerebro pueda interpretar [1].

Una evidencia de este sofisticado procesamiento visual se encuentra en la propia retina. Aunque esta estructura está casi completamente cubierta por conos y bastones —fotorreceptores encargados de captar luz y color—, presenta varios puntos ciegos, como el disco óptico y las venas retinianas (Figura 1.1).

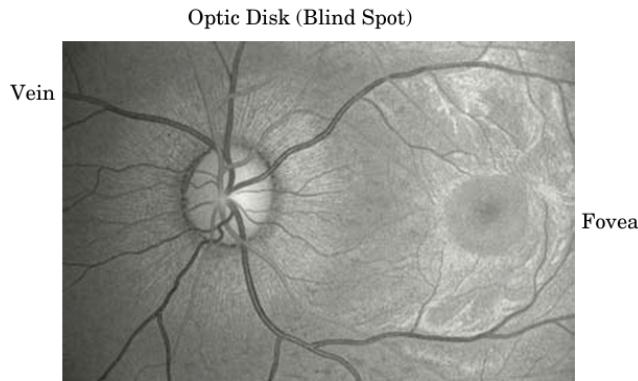


Figura 1.1: Puntos ciegos en la retina. Imagen tomada de [1].

Si tenemos puntos ciegos, ¿por qué entonces nuestra visión no está llena de ellos? La respuesta es que nuestro sistema visual procesa la información de manera que compensa las distorsiones y completa las regiones faltantes de la imagen. Este fenómeno se conoce como “segmentación emergente” y “relleno de características” (*feature filling-in*) [6], y es responsable de ilusiones ópticas como las que se muestran en la Figura 1.2.

La Red de Grossberg busca también replicar el procesamiento visual del brillo. Los seres humanos percibimos el brillo de un objeto siempre estable, aun cuando este varía por cambios globales de luminosidad. Este fenómeno, conocido como “constancia de brillo” [1], nos permite identificar objetos de forma consistente en diferentes contextos de iluminación. Esto sugiere la existencia de un proceso de “normalización de intensidad luminosa” en la corteza visual.

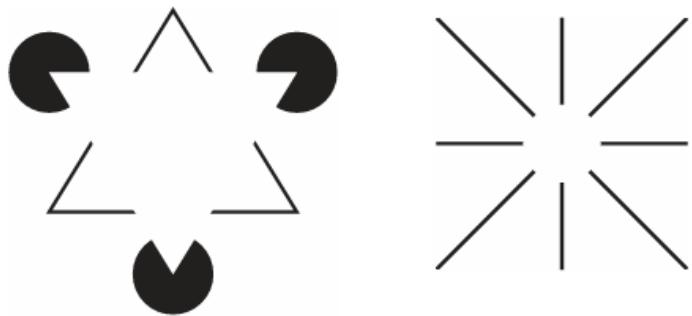


Figura 1.2: Triángulo y círculo ilusorios debido a los procesos del sistema visual. Imagen tomada de [1].

Además, el sistema visual evalúa el brillo de manera comparativa, considerando los gradientes respecto al entorno inmediato. Este fenómeno, denominado “contraste de brillo”, es fundamental para la detección de bordes y formas, por lo que el cerebro está diseñado para identificar y realzar estas diferencias locales [3].

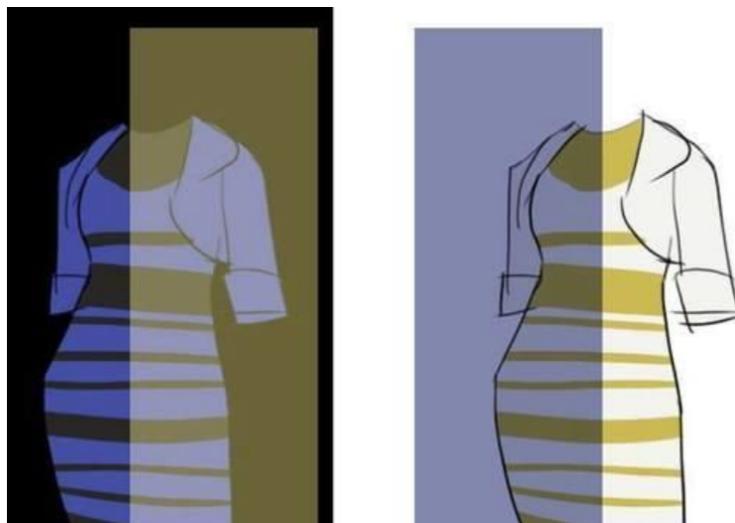


Figura 1.3: Ilusión del vestido azul y negro o blanco y dorado. Imagen tomada de [2].

Sin embargo, este tipo de procesamiento también puede dar lugar a errores perceptivos. Un ejemplo ampliamente conocido es el del denominado “vestido azul y negro o blanco y dorado” (Figura 1.3). En este caso, la ausencia de información clara sobre las condiciones de iluminación en la imagen genera interpretaciones divergentes: algunas personas asumen que el vestido se encuentra en sombra, percibiéndolo como blanco y dorado, mientras que otras lo interpretan como expuesto a luz directa, percibiéndolo como azul y negro.

1.2. Redes Neuronales Recurrentes

Una red neuronal recurrente (RNN, por sus siglas en inglés) es un tipo de red con retroalimentación, lo que significa que los datos de salida se reintroducen como datos de entrada en la red. Por ejemplo, sea $\mathbf{n}(t)$ el vector de activación neuronal en el tiempo t , \mathbf{W} la matriz de pesos sinápticos y \mathbf{b} el vector de sesgos o *bias*; entonces, en el tiempo $t + 1$, la activación de la red recurrente discreta se expresa como [1]:

$$\mathbf{n}(t + 1) = \mathbf{W} \cdot \mathbf{n}(t) + \mathbf{b}. \quad (1.1)$$

Esto permite que la red incorpore el factor temporal en su dinámica de procesamiento, emulando así características fundamentales de los procesos cognitivos biológicos, como la adaptación, la memoria y el aprendizaje. Sin embargo, en muchas aplicaciones no es suficiente conservar únicamente el valor más reciente: se requiere una acumulación progresiva de información a lo largo del tiempo. Para lograrlo, se reemplaza el mecanismo de *step-delay* (1.1) por una integral que actúa sobre todas las entradas previas $\mathbf{p}(t)$, conformando así una memoria extendida.

$$\mathbf{n}(t) = \mathbf{n}(0) + \int_0^t \mathbf{p}(\tau) d\tau. \quad (1.2)$$

No obstante, Grossberg no plantea una acumulación ilimitada de información que pueda saturar las neuronas. La memoria de corto plazo (STM, por sus siglas en inglés) en su modelo es transitoria y decae gradualmente con el tiempo, permitiendo así el procesamiento continuo de nuevos estímulos. Para lograrlo, se emplea una integral con fuga o *leaky integrator*, que incorpora explícitamente un mecanismo de “olvido” [7]:

$$\mathbf{n}(t) = e^{-t/\varepsilon} \mathbf{n}(0) + \frac{1}{\varepsilon} \int_0^t e^{-(t-\tau)/\varepsilon} \mathbf{p}(\tau) d\tau. \quad (1.3)$$

Esta formulación incluye dos componentes: el primero representa el decaimiento u olvido de la respuesta inicial $\mathbf{n}(0)$, mientras que el segundo es una convolución entre el estímulo de entrada $\mathbf{p}(t)$ y una función de decaimiento exponencial. El factor ε controla el grado de retención de información.

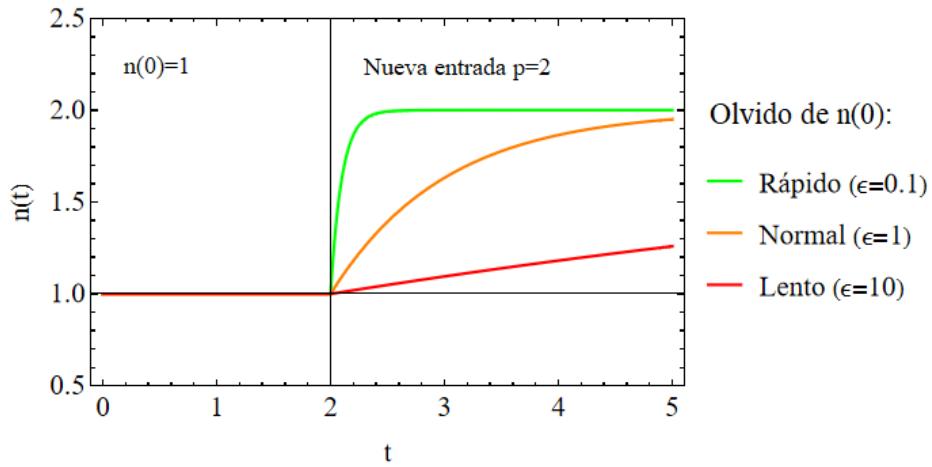


Figura 1.4: *Leaky integrator* $n(0) = 1$, estímulo constante $p = 2$ y distintas velocidades de decaimiento ε .

En la Figura 1.4 se ilustra un escenario donde una condición inicial de $n(0) = 1$ es progresivamente olvidada a favor de una nueva entrada constante $p = 2$, evaluada bajo distintas velocidades de decaimiento ε . El integrador con fuga descrito en (1.3) es solución de la ecuación diferencial:

$$\varepsilon \frac{dn(t)}{dt} = -n(t) + p(t), \quad (1.4)$$

que describe la dinámica de respuesta de una neurona que integra una entrada $p(t)$ mientras su activación decae en el tiempo a un ritmo determinado por ε . Esta ecuación constituye el núcleo del modelo neuronal más fundamental en la Red de Grossberg: el modelo de *Shunting*.

1.3. El Modelo de *Shunting*

El modelo de Shunting es el modelo no lineal que constituye la base fundamental sobre la cual se sostienen los tres mecanismos esenciales en la Red de Grossberg:

- a) **retención de memoria** a corto plazo STM,
- b) **normalización** de la actividad total de la red y
- c) **realce del contraste** entre los patrones de entrada,

todos ellos inspirados por sistema visual humano [3, 4].

A diferencia del *leaky integrator* (1.3), el modelo *Shunting* no solo incorpora el mecanismo STM, sino que además utiliza la recurrencia de la red para normalizar las entradas sin perder la información relativa o contraste entre ellas. De hecho, este modelo no solo conserva el contraste, sino que puede realizarlo activamente [3]. Este comportamiento se logra mediante una comparación dinámica entre entradas diferenciadas como excitatorias $\mathbf{p}^+(t)$ e inhibitorias $\mathbf{p}^-(t)$:

$$\varepsilon \frac{d\mathbf{n}(t)}{dt} = -\mathbf{n}(t) + (\mathbf{b}^+ - \mathbf{n}(t)) \mathbf{p}^+ - (\mathbf{n}(t) + \mathbf{b}^-) \mathbf{p}^-. \quad (1.5)$$

El modelo incorpora además términos de sesgo (*bias*) \mathbf{b}^+ y \mathbf{b}^- , los cuales cumplen un papel central en la normalización de la respuesta, al establecer los límites máximos y mínimos dentro de los cuales la activación neuronal puede oscilar. Estos valores también pueden ajustarse dinámicamente para permitir que el sistema capture y procese amplitudes altamente variables en la señal recibida.

Este trabajo se organiza de la siguiente manera: el capítulo 2 presenta la arquitectura de la Red de Grossberg, incluyendo sus capas funcionales y la regla de aprendizaje. El capítulo 3 introduce la Teoría de Resonancia Adaptativa o ART, detallando la estructura que estabiliza el modelo, el subsistema de orientación y los mecanismos de expectativa y memoria. El capítulo 4 describe los experimentos realizados, comenzando con la aplicación del algoritmo ART-1 en datos artificiales y continuando con los resultados obtenidos en el conjunto MNIST, donde se abordan aspectos como los testores típicos, el problema del sesgo inicial y las estrategias de inicialización supervisada. Finalmente, se presentan las conclusiones del trabajo.

CAPÍTULO 2

RED DE GROSSBERG

La Red de Grossberg está construida a partir de procesos complementarios de memoria a corto y largo plazo, los cuales permiten al sistema adaptarse dinámicamente a los estímulos cambiantes del entorno. Además, esta red incorpora mecanismos fundamentales de segmentación, normalización y realce de contraste, todos ellos inspirados biológicamente y formalizados a través de modelos de Shunting [4].

La arquitectura general de la Red de Grossberg se organiza en dos capas neuronales: *Layer 1* (L1), que representa la retina, y *Layer 2* (L2), que modela la corteza visual junto con los procesos de memoria (STM). Estas capas están interconectadas mediante una matriz de pesos adaptativos, que corresponde a la memoria a largo plazo (LTM), como se ilustra en la Figura 2.1.

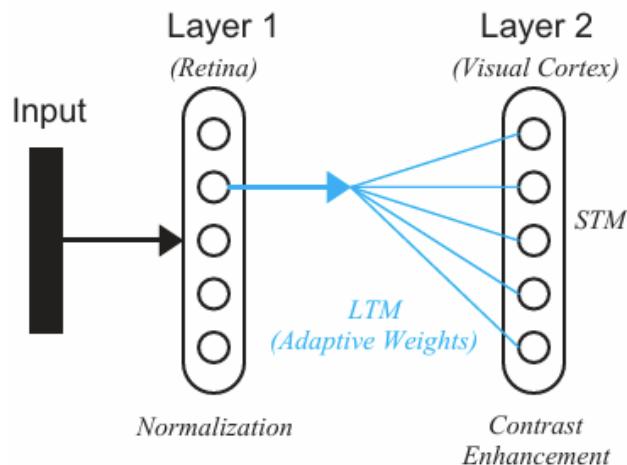


Figura 2.1: Arquitectura de Red de Grossberg. Imagen tomada de [1].

Este diseño dota a la red de plasticidad, permitiéndole adaptarse dinámicamente a las demandas del entorno sin perder la coherencia de los patrones previamente aprendidos.

2.1. Layer 1: Retina

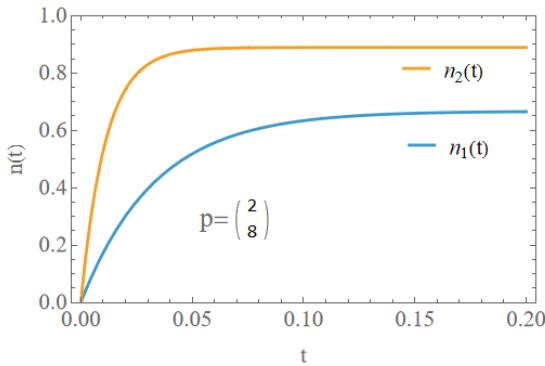
La primera capa es un modelo de *Shunting* que recibe la señal externa y normaliza sus componentes sin perder el contraste relativo entre ellos. Este mecanismo de normalización remite al modo en que nuestra retina procesa las intensidades relativas de la luz en entornos cambiantes, ajustando la percepción para mantener coherencia visual. La ecuación que describe el funcionamiento de esta capa es [1]:

$$\varepsilon \frac{d\mathbf{n}^1(t)}{dt} = -\mathbf{n}^1(t) + ({}^+\mathbf{b}^1 - \mathbf{n}^1(t)) [{}^+\mathbf{W}^1] \mathbf{p} - (\mathbf{n}^1(t) + {}^-\mathbf{b}^1) [\mathbf{W}^1] \mathbf{p}. \quad (2.1)$$

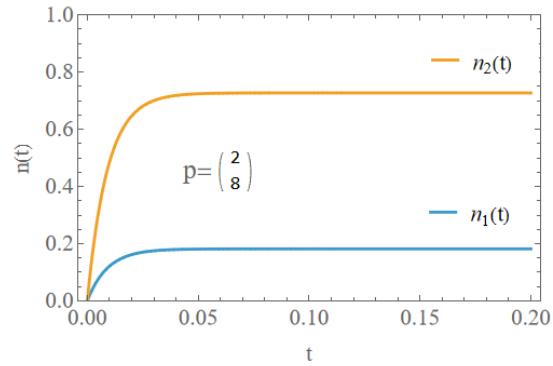
El contraste dinámico se implementa mediante el método On-Center, Off-Surround, en el cual la entrada excitatoria de la neurona i se localiza en la posición i (centro), mientras que las posiciones circundantes ejercen una influencia inhibidora sobre ella. Este mecanismo se modela matemáticamente a través de matrices que simulan el comportamiento de las células ganglionares en la retina [3]:

$${}^+\mathbf{W}^1 = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{W}^1 = \begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 0 \end{bmatrix}. \quad (2.2)$$

El parámetro ε determina la rapidez de respuesta del sistema, y se elige de forma que la activación neuronal se estabilice antes de que ocurran cambios en los pesos adaptativos del modelo.



(a) Normalización de *leaky integrator*



(b) Normalización de *Shunting model*

Considérese la normalización ${}^+b = 1$ aplicada a una entrada $\mathbf{p} = [2, 8]$, procesada por dos neuronas n_1 y n_2 , como se muestra en la Figura (a) para el modelo de leaky integrator y en la Figura (b) para el modelo de Shunting. Mientras en (a) el proceso de normalización pierde el contraste original entre los valores 2 y 8, en (b) la diferencia relativa entre ellos se preserva.

2.2. Layer 2: Corteza Visual y STM

La segunda capa recibe el *output* \mathbf{a}^1 de la primera capa L1, el cual viene multiplicado por los pesos sinápticos \mathbf{W} de LTM. La capa L2 también se modela mediante un esquema de tipo *Shunting*. Sin embargo, a diferencia de la capa anterior, la función que cumple es el realce de contraste y memoria STM de la corteza visual [3].

Las neuronas en esta capa representan a los clústers o clases, y compiten entre sí mediante el modelo de Shunting. Esta competencia da lugar a un mecanismo de selección, en el que la clase ganadora se almacena temporalmente en el sistema, funcionando como una memoria a corto plazo STM. La ecuación que describe el funcionamiento de esta capa es [1]:

$$\varepsilon \frac{d\mathbf{n}^2(t)}{dt} = -\mathbf{n}^2(t) + ({}^+b^2 - \mathbf{n}^2(t)) \{ [{}^+W^2] f^2(\mathbf{n}^2(t)) + \mathbf{W}^2 \mathbf{a}^1 \} - (\mathbf{n}^2(t) + {}^-b^2) [{}^-W^2] f^2(\mathbf{n}^2(t)) \quad (2.3)$$

El realce de contraste y la competencia entre neuronas están también mediados por el método *On-Center, Off-Surround* y por una función de activación f . El tipo de función elegida determina el comportamiento del sistema: una función f lineal permite almacenar cualquier patrón con fidelidad, pero también amplifica el ruido; una función sublineal, en cambio, tiende a amplificar únicamente el ruido y reduce el contraste; una función supralineal genera un efecto de *winner-take-all*, favoreciendo que una sola clase domine; finalmente, una función sigmoide logra equilibrar estos aspectos, ya que suprime el ruido, mejora el contraste y evita una cuantización extrema [1].

Para ilustrar el funcionamiento del sistema, la Figura 2.2 muestra la respuesta de dos neuronas

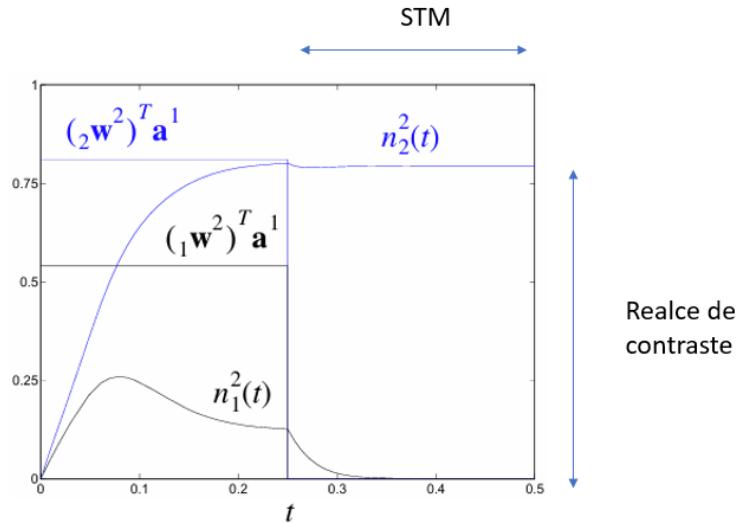


Figura 2.2: Competencia de dos neuronas en L2. Imagen adaptada de [1].

ante entradas de 0.54 y 0.81, respectivamente. Se observa que el contraste entre ambos valores se incrementa, evidenciando el efecto de realce generado por la competencia neuronal. Además, el patrón de activación resultante se mantiene incluso después de retirar los estímulos de entrada, lo que demuestra la capacidad del modelo para retener información de manera transitoria, funcionando como una memoria a corto plazo (STM).

2.3. Pesos adaptativos: LTM

Finalmente, respecto a los pesos sinápticos, Grossberg se refiere a ellos como pesos adaptativos o memoria a largo plazo LTM (del inglés *Long Term Memory*), ya que las filas de la matriz \mathbf{W} representan los patrones almacenados que la red es capaz de reconocer. La ley de aprendizaje utilizada es una versión de la regla de Hebb con decaimiento [1]:

$$\frac{d[i\mathbf{W}^2(t)]}{dt} = \alpha \left\{ -[i\mathbf{W}^2(t)] + n_i^2(t)\mathbf{n}^1(t) \right\}. \quad (2.4)$$

El primer término es el de decaimiento y evita la saturación de LTM. El segundo es un término de aprendizaje asociativo, basado en el principio de Hebb: “las neuronas que se activan juntas, se conectan juntas”.

Sin embargo, una limitación de la regla de Hebb con decaimiento es que requiere de estímulos constantes; de lo contrario, las asociaciones previamente formadas tienden a desaparecer. En una formulación alternativa conocida como *Instar Learning Rule* el peso decae únicamente cuando su neurona ha sido activada. Este enfoque evita penalizar conexiones inactivas $n_i = 0$, concentrando el ajuste sináptico en aquellas neuronas que realmente participan en la codificación del estímulo.

$$\frac{d [{}_i \mathbf{W}^2(t)]}{dt} = \alpha \cdot n_i^2(t) \{ - [{}_i \mathbf{W}^2(t)] + \mathbf{n}^1(t) \} \quad (2.5)$$

De esta forma, la arquitectura de la Red de Grossberg, con la base dinámica del modelo de *Shunting*, logra capturar las propiedades experimentales observadas en el sistema visual humano: plasticidad (recodificación universal), *feature fill-in*, STM, LTM, normalización y realce de contraste. Sin embargo, Grossberg se dio cuenta que su red tiene un problema de estabilidad en la formación de clústers debido a esa gran plasticidad.

Grossberg demostró [5] que, si el número de patrones de entrada no es demasiado grande o si dichos patrones no forman demasiados clústeres en relación con el número de neuronas en L2, el aprendizaje eventualmente se estabiliza. Sin embargo, también mostró que las redes competitivas estándar no presentan un aprendizaje estable cuando se enfrentan a flujos de patrones de entrada arbitrarios. Este hallazgo lo llevó a implementar el mecanismo de resonancia adaptativa ART (del inglés *Adaptive Resonance Theory*), diseñado para estabilizar la plasticidad de la red.

CAPÍTULO 3

ADAPTATIVE RESONANCE THEORY

¿Cómo puede un sistema ser receptivo a nuevos patrones y, al mismo tiempo, mantener patrones previos? Sabemos que los sistemas biológicos son muy eficaces en esto. Por ejemplo, podemos reconocer fácilmente el rostro de nuestra madre, incluso si ha transcurrido mucho tiempo y hemos conocido muchos rostros nuevos en el camino.

El sistema olfativo proporciona a Grossberg la motivación perfecta para la modelo ART [5]. En concreto, la interacción entre el bulbo olfativo y la corteza olfativa. El elemento clave de esta inspiración es la capacidad de la corteza olfativa de formar patrones que retroalimentan al bulbo como “expectativas”. Si el patrón olfativo que llega del bulbo coincide con la expectativa de la corteza, la corteza acepta dicha activación. De lo contrario, el patrón se extingue sin desestabilizar el clúster de olores. La arquitectura de una Red ART (Figura 3.1) parte de la

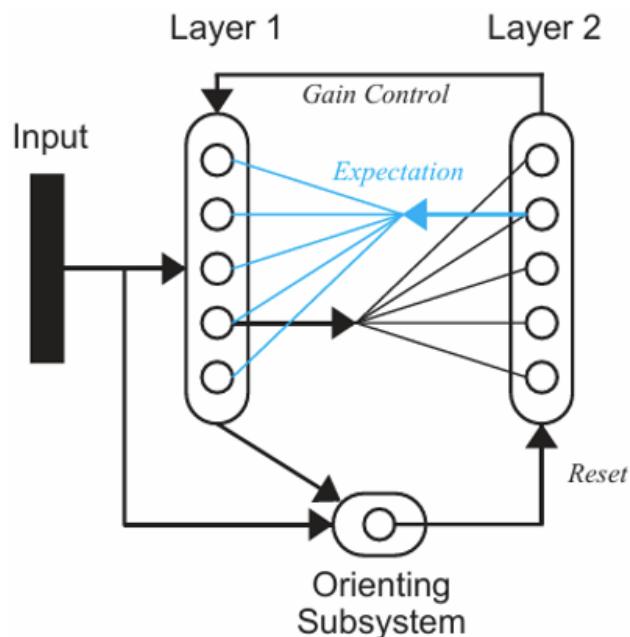


Figura 3.1: Arquitectura de Red ART. Imagen tomada de [1].

Red de Grossberg como base. La principal diferencia es la implementación del sistema de expectativas para cada clúster, el cual determina cuándo una nueva entrada es lo suficientemente compatible como para modificar ese clúster.

3.1. Layer 1: Bulbo Olfativo

En la Red ART, la capa L1 ya no se encarga de normalizar las señales de entrada; en este modelo, dicha tarea es delegada a la regla de aprendizaje que ajusta los pesos $\mathbf{W}^{1:2}$. En su lugar, L1 implementa un mecanismo de comparación entre el patrón recibido (señal externa \mathbf{p}) y el patrón esperado (experctativa del clúster ganador en L2 $\mathbf{W}^{2:1}\mathbf{a}^2$). La ecuación se operación sigue también un modelo de *Shunting* [8]:

$$\varepsilon \frac{d\mathbf{n}^1(t)}{dt} = -\mathbf{n}^1(t) + ({}^+\mathbf{b}^1 - \mathbf{n}^1(t)) \{ \mathbf{p} + \mathbf{W}^{2:1}\mathbf{a}^2(t) \} - (\mathbf{n}^1(t) + {}^-\mathbf{b}^1) [{}^-\mathbf{W}^1] \mathbf{a}^2(t), \quad (3.1)$$

donde la entrada excitatoria contiene ambos patrones a comparar \mathbf{p} y $\mathbf{W}^{1:2}\mathbf{a}^2$. La entrada inhibitoria, por otro lado, es un término de control de ganancia con matriz ${}^-\mathbf{W}^1$

$${}^-\mathbf{W}^1 = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}. \quad (3.2)$$

El propósito de este término es regular la activación neuronal y, en la práctica, es la diferencia entre un operador OR o AND en la comparación entre \mathbf{p} y $\mathbf{W}^{1:2}\mathbf{a}^2$.

Realizar un análisis en estado estacionario (es decir, cuando $\frac{dn}{dt} = 0$) o en el régimen de aprendizaje rápido ($\varepsilon \ll 1$) resulta particularmente útil para comprender el funcionamiento interno de los mecanismos que conforman una Red ART. Además, asumir *inputs* binarios simplifica considerablemente las ecuaciones. El análisis estacionario de la ecuación 3.1 muestra

que [1]:

$$\mathbf{a}^1 = \begin{cases} \mathbf{p}, & \text{si } a_j^2 = 0 \text{ para todo } j, \\ \mathbf{p} \cap \mathbf{W}_j^{2:1}, & \text{si } a_j^2 = 1 \text{ para un } j, \end{cases}$$

donde \mathbf{a}_1 el *output* de L1 corresponde a la operación (\mathbf{p} AND $\mathbf{W}_j^{2:1}$), y al ser vectores binarios la norma $\|\mathbf{a}^1\|$ funciona como un contador de coincidencias entre el patrón entrante y el patrón esperado.

3.2. Layer 2: Corteza Olfativa y STM

Por otro lado, la capa L2 en la Red ART es la misma que la capa L2 en la Red Grossberg. Es decir, un modelo de *Shunting* con ecuación 2.3 que promueve la competencia neuronal y almacena la clase ganadora en la STM.

En el límite de aprendizaje rápido y para entradas binarias, al igual que ocurre en L1, el funcionamiento de L2 se simplifica a una operación de competencia básica [1]:

$$a_i^2 = \begin{cases} 1, & \text{si } \left(\left({}_i \mathbf{W}^{1:2} \right)^T \mathbf{a}^1 = \max \left[\left(\mathbf{W}^{1:2} \right)^T \mathbf{a}^1 \right] \right) \\ 0, & \text{otro} \end{cases} \quad (3.3)$$

Es decir, la neurona con mayor entrada en L2 gana, mientras que el resto de neuronas o clústers quedan suprimidos, siguiendo un esquema de *winner-take-all*. Una vez seleccionada la neurona o clúster ganador en L2, este proyecta su patrón expectativa almacenado en $\mathbf{W}^{2:1}$ de vuelta a L1, donde se compara con el patrón de entrada.

3.3. Subsistema de Orientación

Uno de los elementos clave de la arquitectura ART es el Subsistema de Orientación a^0 . Su propósito es determinar si existe o no una coincidencia suficiente entre el patrón de expectativa

$\mathbf{W}_j^{2:1}$, proyectado de L2 a L1 por el clúster ganador j , y su patrón de entrada \mathbf{p} [1].

$$a^0 = \begin{cases} 1, & \text{si } \left[\|\mathbf{a}^1\|^2 / \|\mathbf{p}\|^2 < \rho \right] \\ 0, & \text{otro} \end{cases} \quad (3.4)$$

donde $\mathbf{a}^1 = \mathbf{p} \cap \mathbf{W}_j^{2:1}$ y $\|\mathbf{a}^1\|^2$ es el número de coincidencias entre patrones \mathbf{p} y $\mathbf{W}_j^{2:1}$ mientras $\|\mathbf{p}\|^2$ normaliza la expresión. A partir de ello, se define un parámetro de vigilancia $\rho \in [0, 1]$, que controla el nivel de exigencia requerido para aceptar una coincidencia como válida. Valores altos de ρ implican una mayor rigurosidad, lo que se traduce en una clasificación más fina y una mayor granularidad en la formación de clústeres.

Por ejemplo, si $\|\mathbf{a}^1\|^2 / \|\mathbf{p}\|^2 < \rho$, no hubo coincidencia suficiente. Esto implica que se debe inhibir el clúster ganador y buscar una coincidencia en el segundo cluster más cercano y así hasta superar el umbral. Cuando $\|\mathbf{a}^1\|^2 / \|\mathbf{p}\|^2 \geq \rho$ hubo una coincidencia y procedemos a actualizar ambas matrices de pesos $\mathbf{W}^{1:2}$ y $\mathbf{W}^{2:1}$ para incluir el nuevo *input* en un mecanismo conocido como “resonancia”.

3.4. Pesos adaptativos: Expectativa y LTM

El mayor cambio de la Red ART con respecto a la Red de Grossberg es la inclusión de una matriz de expectativas $\mathbf{W}^{2:1}$. Esta matriz almacena los patrones esperados asociados a cada clúster en L2, permitiendo compararlos con los nuevos patrones de entrada en L1 para determinar si estos son lo suficientemente compatibles.

Similar a la ley de aprendizaje de la Red de Grossberg, $\mathbf{W}^{2:1}$ se actualiza mediante una regla hebbiana con decaimiento. La única diferencia está en el cambio de sentido (de L2 a L1). Por esta razón, la regla de aprendizaje pasa a denominarse *Outstar Learning Rule*, ya que sigue dependiendo de la activación del clúster ganador en L2.

$$\frac{d [\mathbf{W}_j^{2:1}(t)]}{dt} = a_j^2(t) \left[-\mathbf{W}_j^{2:1}(t) + \mathbf{a}^1(t) \right]. \quad (3.5)$$

Por otra parte, los pesos $\mathbf{W}^{1:2}$ de L1 a L2, utilizan un tipo de *Instar Learning Rule* junto con un modelo de *Shunting On-Center, Off-Surround* que regulariza $\mathbf{W}^{1:2}$ y evita dilemas de subset/superset. La ecuación de este modelo es [1]:

$$\begin{aligned} \frac{d [{}_i \mathbf{W}^{1:2}(t)]}{dt} &= a_i^2(t) \left[\{ {}^+ \mathbf{b} - {}_i \mathbf{W}^{1:2}(t) \} \zeta [{}^+ \mathbf{W}] \mathbf{a}^1(t) \right. \\ &\quad \left. - \{ {}_i \mathbf{W}^{1:2}(t) + {}^- \mathbf{b} \} [{}^- \mathbf{W}] \mathbf{a}^1(t) \right], \end{aligned} \quad (3.6)$$

aunque en el régimen de aprendizaje rápido se simplifica. Este régimen implica que el *output* de L2 (STM) permanece constante frente a las variaciones de $\mathbf{W}^{1:2}(t)$ y $\mathbf{W}^{2:1}(t)$ (LTM) durante el aprendizaje.

$$\mathbf{W}_j^{2:1} = \mathbf{a}^1 \quad (3.7)$$

$${}_j \mathbf{W}^{1:2} = \frac{\zeta \mathbf{a}^1}{\zeta + \|\mathbf{a}^1\|^2 - 1} \quad (\text{regularizado}). \quad (3.8)$$

Estas ecuaciones representan la actualización de pesos para el clúster ganador j , donde $\zeta > 1$ es la velocidad de aprendizaje y $\mathbf{a}^1 = \mathbf{p} \cap \mathbf{W}_j^{2:1}$ el patrón que incorpora LTM de la red. De este modo, la red incorpora únicamente las coincidencias entre las expectativas generadas y los nuevos estímulos, reforzando lo aprendido y manteniendo la coherencia de cada categoría.

En el siguiente capítulo se abordará el modelo ART-1, una versión simplificada del modelo ART diseñada para entradas binarias y régimen de aprendizaje rápido. Este modelo ha sido elegido para su análisis no solo por su probada estabilidad [8], sino también porque su algoritmo se simplifica notablemente bajo estas condiciones —como se ha demostrado a lo largo de este capítulo—, lo que lo convierte en una herramienta ideal para aplicaciones prácticas de reconocimiento de patrones.

CAPÍTULO 4

EXPERIMENTOS

Para validar el funcionamiento del modelo ART-1 (aprendizaje rápido con *inputs* binarios), se realizó un experimento utilizando un conjunto de datos artificiales compuesto por representaciones binarizadas de las letras A, B, C y D en una cuadrícula de 7x6 píxeles. Los parámetros del modelo fueron definidos como sigue: una tasa de aprendizaje $\zeta = 2$, parámetro de vigilancia ρ variable, y pesos sinápticos inicializados uniformemente. La matriz de expectativas $\mathbf{W}^{2:1}$ se inicializa toda a 1s mientras la matriz de pesos $\mathbf{W}^{1:2}$ se regulariza como indica la ecuación 3.8.

4.1. Algoritmo ART-1

El procedimiento se desarrolla en seis pasos [1]:

1. Se presenta un patrón de entrada \mathbf{p}_1 a la red. Dado que la L2 está inactiva en la inicialización, se asigna $\mathbf{a}^1 = \mathbf{p}_1$ en concordancia con 3.1.
2. Se calcula el *input* de L2 multiplicando $\mathbf{W}^{1:2}$ por \mathbf{a}^1 , y se activa la neurona con mayor valor como indica 3.3. En este caso, de empate la neurona ganadora es la del clúster 1 (se toma en orden).
3. Se genera la expectativa de la red proyectando \mathbf{a}^2 desde L2 a L1, es decir, $\mathbf{W}^{2:1}\mathbf{a}^2$. Esta expectativa se compara con el patrón externo mediante la intersección lógica $\mathbf{a}^1 = \mathbf{p}_1 \cap \mathbf{W}_j^{2:1}$ en concordancia con 3.1.
4. Se verifica la condición de resonancia evaluando de acuerdo con 3.4. Si

$$\frac{\|\mathbf{a}^1\|^2}{\|\mathbf{p}_1\|^2} \geq \rho$$

pasamos al paso 5. Caso contrario, se inhibe el clúster ganador j y se repetiría el proceso desde el paso 1 con el siguiente clúster más activo.

5. Resonancia. Se actualizan los pesos de la red de acuerdo a 3.8:

$$\mathbf{W}_j^{1:2} = \frac{\zeta \mathbf{a}^1}{\zeta + \|\mathbf{a}^1\|^2 - 1}, \quad {}_j \mathbf{W}^{2:1} = \mathbf{a}^1,$$

reforzando únicamente las coincidencias entre la entrada y la expectativa.

6. Finalmente, se restauran las neuronas inhibidas en L2 y se ingresa un nuevo patrón.

4.2. Resultados en Datos Artificiales

El parámetro de vigilancia ρ define el umbral mínimo de coincidencia requerido entre un nuevo patrón de entrada y una categoría almacenada para consolidar el aprendizaje. La Figura

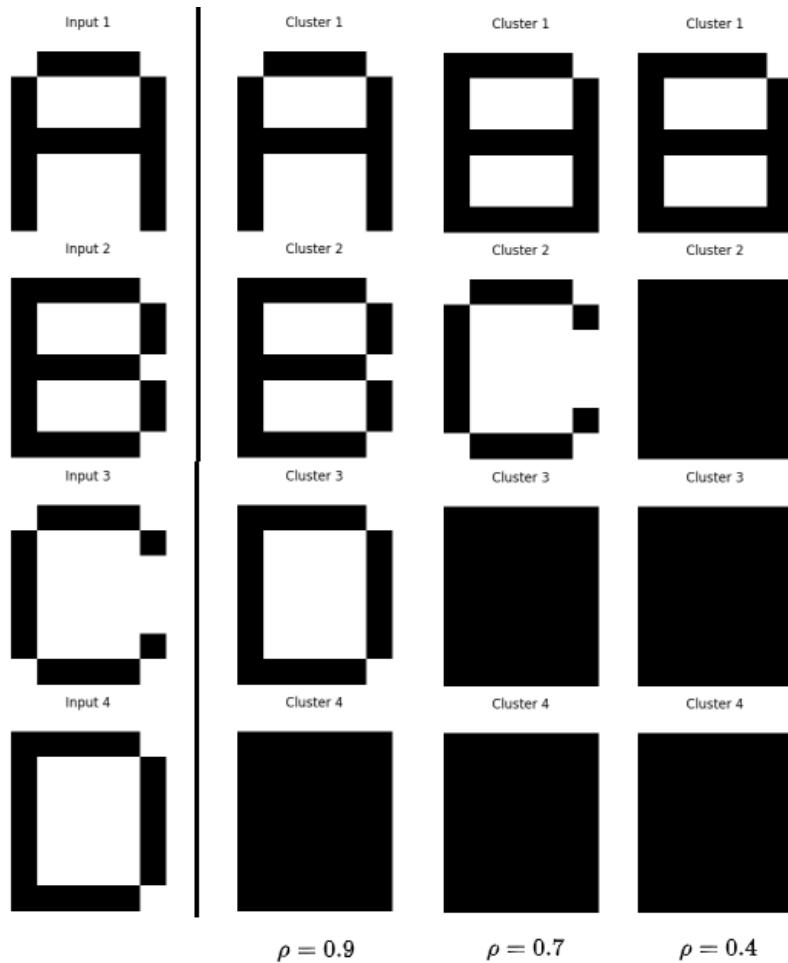


Figura 4.1: Letras A, B, C y D clasificadas con parámetros de vigilancia $\rho = 0.9, 0.7$ y 0.4 .

4.1 permite observar con claridad cómo el parámetro de vigilancia influye en la agrupación

de patrones con diferentes niveles de similitud. Con un valor alto del parámetro de vigilancia (ρ cercano a 1), el modelo distingue entre todas las letras, asignando una clase distinta a cada patrón binario, incluso entre las letras C y D, cuyas representaciones gráficas comparten varias características estructurales. Sin embargo, al reducir el valor de ρ , se observó que C y D fueron agrupadas dentro de una misma categoría, así también A y B; y finalmente, todas terminan dentro de una misma clase.

Este comportamiento pone en evidencia la flexibilidad del modelo ART-1 para ajustar la granularidad del agrupamiento según se requiera. Sin embargo, dicho ajuste conlleva a la formación de clústers menos precisos o menos coherentes internamente, afectando potencialmente la calidad y pureza general de la clasificación.

4.3. Resultados en MNIST

Para propósitos reales de prueba, se emplea el conjunto de datos MNIST, ampliamente utilizado en modelos de clasificación y predicción. MNIST es un conjunto de dígitos manuscritos almacenados en matrices de 28x28 píxeles. La librería dispone de 60,000 imágenes para entrenamiento y 10,000 para prueba [9].

Antes de aplicar ART-1, es necesario convertir las imágenes de dígitos MNIST —originalmente en escala de grises en el rango [0, 255]— a un formato binario, ya que ART-1 opera únicamente con entradas binarias. Para ello, se emplea un umbral de 100, tal como se recomienda en [10], de modo que cualquier valor superior se asigna como 1, y los valores iguales o inferiores se asignan como 0.

El modelo ART-1 fue aplicado a un subconjunto aleatorio de 350 imágenes binarizadas de MNIST, utilizando un parámetro de vigilancia $\rho = 0.9$, alcanzando una precisión (medida como “pureza” de los clusters) del 98 %. Sin embargo, para este ρ de 0.9 se generaron 334 clases distintas!, un resultado también reportado por [11]. Esto evidencia una severa sobresegmentación

del espacio de 10 clases reales.

Para mitigar esta fragmentación excesiva, el parámetro de vigilancia fue reducido a un valor más flexible de $\rho = 0.01$, lo que permitió disminuir el número de clases detectadas a 10. No obstante, esta estrategia tuvo un impacto negativo significativo en la precisión del modelo, que descendió a apenas 32 %.

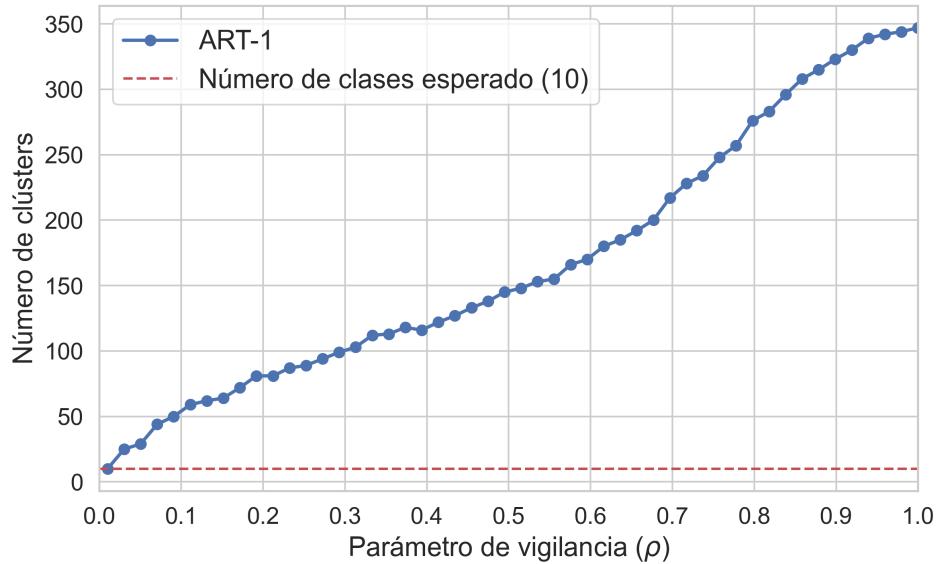


Figura 4.2: Distribución del número de clases generadas por ART-1 para varios ρ en el subconjunto MNIST de 350 imágenes.

Este resultado, ilustrado en la figura 4.2, señala que para una buena clasificación de MNIST, ART-1 requiere una sobresegmentación significativa (de 334 clases frente a las 10 clases reales del conjunto). MNIST es un conjunto de datos complejo y de alta dimensionalidad que desafía la capacidad adaptativa (plasticidad) del modelo. Esta situación pone de manifiesto la necesidad de aplicar técnicas de reducción de atributos que permitan disminuir la dimensionalidad y complejidad del espacio de representación y así mejorar la estabilidad y eficiencia del proceso de clasificación.

4.3.1. Testores Típicos

Con este fin se implementó el algoritmo de preprocessamiento MinReduct [12], diseñado para calcular los reductos más cortos en un sistema de decisión, es decir, los subconjuntos mínimos de atributos que preservan la capacidad de clasificación del conjunto completo. Esta combinación permite reducir significativamente el tamaño de los datos y el número de atributos sin perder información relevante, optimizando así el rendimiento y la velocidad del modelo ART-1 aplicado a MNIST.

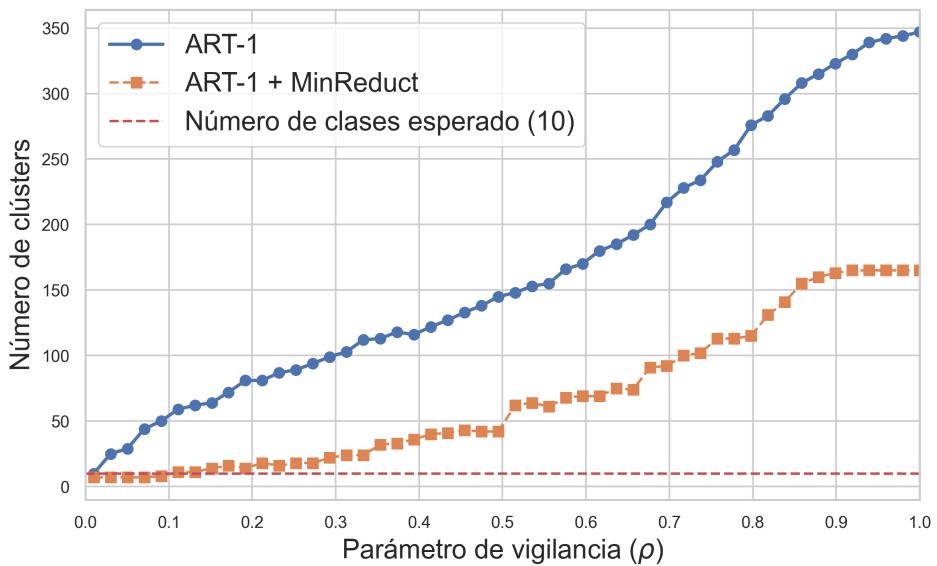


Figura 4.3: Distribución del número de clases generadas por ART-1 para varios ρ en el subconjunto MNIST de 350 imágenes.

La aplicación del algoritmo MinReduct redujo el conjunto de características al 19,38 % del total original, logrando disminuir el tiempo promedio de entrenamiento de 2.41 segundos a 0.63 segundos. Esto corresponde a una mejora del 73.9 % en eficiencia temporal, que será de gran relevancia cuando entrenemos toda la base de 60.000 ejemplares. Esta compactación del espacio de características también se tradujo en una mejora notable en la estabilidad del modelo, así como en una disminución significativa del número de clústers requeridos para alcanzar una calidad de $\rho = 0.9$. Sin embargo, el espacio de clases sigue estando significativamente sobresegmentado (163 clases frente a las 10 clases reales).

4.3.2. El Problema del Sesgo Inicial

Si forzamos al modelo a generar exactamente las 10 clases que hay en MNIST surge un problema, se observa un importante sesgo hacia los 10 primeros números aleatorios en el conjunto de datos. Por ejemplo, supongamos que estos son los que se muestran en la Figura 4.4.

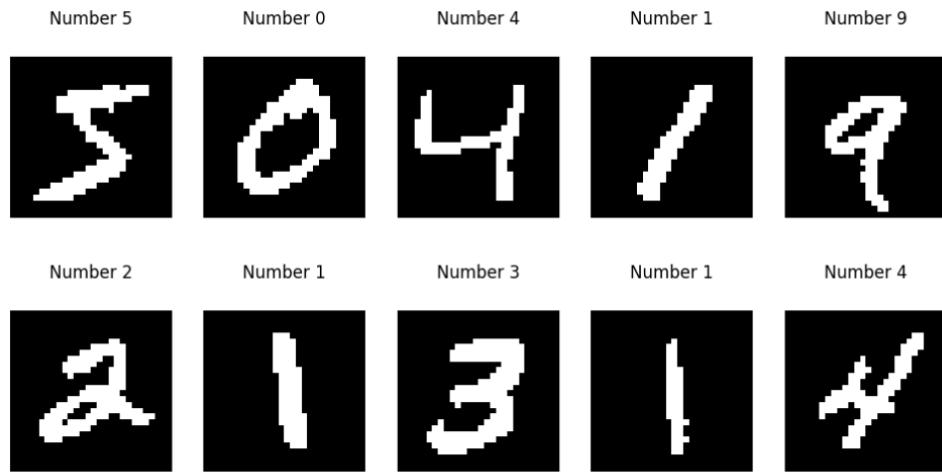


Figura 4.4: Primeros 10 inputs de MNIST sobre los que entrena el modelo ART-1

Como resultado del entrenamiento completo sobre los 60,000 ejemplos disponibles en MNIST, los patrones $\mathbf{W}^{2:1}$ aprendidos y almacenados en la memoria a largo plazo LTM para cada clúster son los que se presentan en la Figura 4.5.

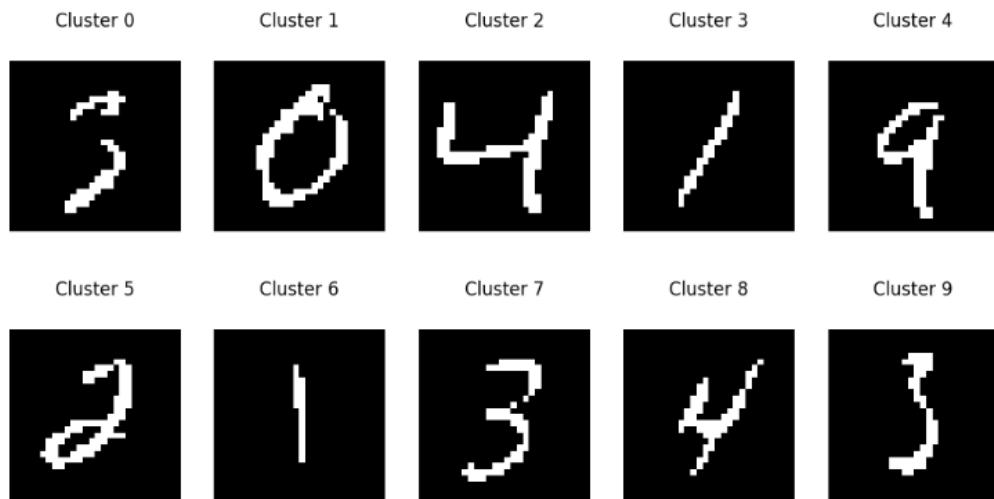


Figura 4.5: Patrones $\mathbf{W}^{2:1}$ almacenados en la LTM del modelo ART-1 después de entrenar todo MNIST

Aun bajo un valor relativamente flexible del parámetro de vigilancia $\rho = 0.8$ se puede apre-

ciar la fuerte influencia de los primeros diez patrones presentados al modelo. En el experimento realizado, los dígitos 6, 7 y 8 no estuvieron representados entre las primeras muestras aleatorias. Como consecuencia, y pese a que el modelo fue entrenado posteriormente con el conjunto completo de 60.000 imágenes, estos dígitos nunca llegaron a ser entrenados ni forman parte de la memoria a largo plazo LTM del modelo. En otras palabras, los dígitos 6, 7 y 8 no se clasificaron, situando la precisión de clasificación de ART-1 en un bajo 44 %.

En particular, cuando los primeros patrones no representan de manera equitativa todas las clases reales del conjunto —como ocurre en escenarios de inicialización aleatoria no balanceada—, el modelo excluye permanentemente aquellas clases ausentes, especialmente bajo valores elevados de ρ . La estabilidad que caracteriza a ART-1, si bien valiosa para evitar la catástrofe del olvido, también limita su capacidad de corrección posterior, generando una segmentación parcial del espacio de clases que impacta negativamente la calidad de clasificación y precisión del modelo.

4.3.3. Estrategias de Inicialización Supervisada

Para mitigar este problema, se diseñó una estrategia que consiste en dirigir la inicialización de las matrices $\mathbf{W}^{1:2}$ y $\mathbf{W}^{2:1}$ de memoria a largo plazo (LTM) —las cuales hasta el momento se inicializaban uniformemente en 1— para que ahora representen explícitamente cada una de las 10 clases verdaderas del conjunto MNIST (0–9). Esta elección supervisada garantiza que cada clase esté representada desde el inicio del entrenamiento, evitando así el problema del sesgo inicial.

Para seleccionar patrones representativos de cada una de las 10 clases verdaderas, puede hacerse simplemente de manera aleatoria. Esta estrategia ya resuelve el problema del sesgo inicial previamente discutido, gracias a lo cual la precisión mejora ligeramente del 44 % al 48.29 %. Sin embargo, la selección aleatoria puede asignar malos representantes, que no capturan adecuadamente las características generales de su clase. Una estrategia más eficaz consiste en

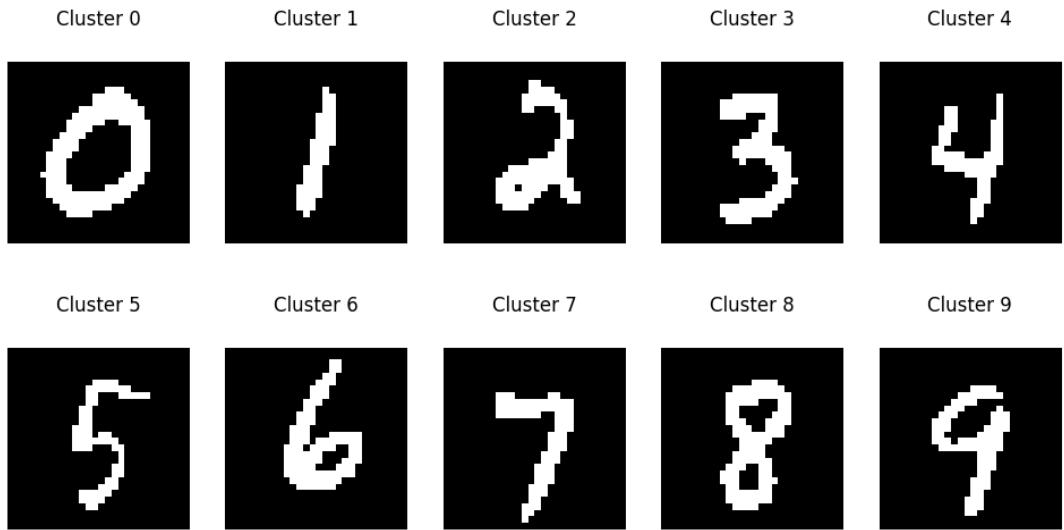


Figura 4.6: Pesos iniciales $\mathbf{W}^{1:2}$, $\mathbf{W}^{2:1}$ más cercanos al centroide de cada clase.

elección como patrones de inicialización aquellos ejemplos cercanos al centroide de cada conjunto de clase, ya que representan de forma más fiel al grupo, como se ilustra en la Figura 4.6. Al implementar esta estrategia, la precisión del modelo mejora significativamente, alcanzando un 62.27 %.

Sin embargo, la estrategia más eficaz resulta ser una combinación de ambos enfoques: además de utilizar el mejor representante (es decir, el ejemplo más cercano al centroide de cada clase), se incorporan también patrones aleatorios que capturan la diversidad de formas en que un mismo número puede ser escrito. Lo que se hace es permitir múltiples prototipos por clase (por ejemplo, cinco patrones representativos en lugar de uno solo). Con esta aproximación, la precisión del modelo mejora a un 69.21 %. Al incrementar aún más el número de representantes a diez, se alcanza la mayor precisión registrada para este modelo, con un 78 %.

Si bien es posible seguir añadiendo representantes aleatorios para intentar mejorar la precisión, esto reintroduce los problemas de sobreajuste y sobresegmentación. Los resultados obtenidos se resumen a continuación en la siguiente tabla 4.1.

Cuadro 4.1: Precisión alcanzada por el modelo ART-1 bajo diferentes estrategias de inicialización.

Tipo de inicialización	Precisión alcanzada (%)
Sin inicialización dirigida (aleatoriedad total)	44.00
Inicialización aleatoria por clase	48.29
Inicialización por centroide por clase	62.27
Centroide + 4 patrones aleatorios por clase	69.21
Centroide + 9 patrones aleatorios por clase (mayor precisión)	78.00

4.3.4. Comparación con Métodos de Referencia

Si bien la precisión alcanzada por el modelo ART-1 no resulta destacable al compararse con los modelos que actualmente son referentes en la tarea de clasificación del conjunto MNIST —presentados en la Tabla 4.2—, es importante resaltar que ART-1 pertenece a una clase distinta de modelos. A diferencia de los enfoques supervisados y estáticos, ART-1 opera bajo un paradigma no supervisado y dinámico, lo que le confiere propiedades valiosas que los modelos referentes no poseen.

Cuadro 4.2: Comparación de precisión reportada en MNIST por modelos supervisados y estáticos.

Modelo de clasificador	Precisión reportada (%)
Clasificador lineal (1-capa NN) [13]	88.00
K-nearest neighbors, Euclidiano (L2) [13]	95.00
40 PCA + clasificador cuadrático [13]	96.70
SVM, kernel Gaussiano [14]	98.60
Red neuronal 2 capas, 800 HU, pérdida cross-entropy [15]	98.40
Red neuronal 3 capas, 500 + 300 HU, softmax, cross-entropy, weight decay [14]	98.47

Por ejemplo, ART-1 es un modelo especialmente adecuado para escenarios de aprendizaje realista a gran escala, donde no se dispone de etiquetas previas ni resulta viable reentrenar completamente una red cada vez que se incorpora nueva información. Gracias a su arquitectura, ART-1 es capaz de aprender de forma rápida y continua, adaptándose a la creciente y constante llegada y acumulación de datos, sin requerir procesos de reentrenamiento. Esto lo diferencia de los modelos supervisados estáticos, que, si bien superan a ART-1 en precisión, presentan limitaciones cuando se trata de manejar flujos dinámicos de información en tiempo real.

CAPÍTULO 5

CONCLUSIONES

A pesar de que el modelo ART-1 presentado en este trabajo alcanzó una precisión del 78 % en la tarea de clasificación del conjunto MNIST —cifra menor en comparación con modelos de referencia actuales—, su aplicación sigue siendo de gran valor en aplicaciones a gran escala. El modelo ART-1, al estar inspirado en el funcionamiento biológico del sistema cognitivo humano, responde a principios de aprendizaje no supervisado, dinámico e incremental.

Esto significa que, a diferencia de las arquitecturas estáticas, que requieren reentrenamiento completo para incorporar nuevos patrones —proceso costoso en términos computacionales—, ART-1 es capaz de aprender en tiempo real, y sin sufrir del *catastrophic forgetting* que afecta a modelos de *deep learning* [16]. Por tanto, ART-1 ofrece un enfoque especialmente valioso para enfrentar las demandas del contexto tecnológico actual, donde los desafíos relacionados con el entrenamiento costoso y la actualización continua de modelos son cada vez más relevantes.

Además, es posible mejorar el rendimiento del modelo —o, más precisamente, facilitarle la comprensión del conjunto de datos— mediante técnicas de reducción de dimensionalidad, como el algoritmo MinReduct, PCA [17] o convoluciones [11]. A esto puede sumarse una estrategia de ajuste semisupervisado de los parámetros iniciales, que guía al modelo en la conformación de clústers más coherentes desde el inicio del entrenamiento. La simplicidad inherente del modelo ART-1 permite que una amplia variedad de técnicas [18–24] puedan aplicarse para adaptarlo de manera eficaz a bases de datos complejas.

En este sentido, es relevante señalar que existen extensiones más robustas dentro de la familia ART que podrían superar las limitaciones de ART-1 en MNIST discutidas en este trabajo.

Modelos como ART-2 [25, 26], capaz de operar sobre datos continuos, o ARTMAP [27, 28], que incorpora supervisión explícita en el proceso de aprendizaje manteniendo el dinamismo característico, han demostrado desempeños notables. De hecho, ARTMAP, al integrar lógica difusa o *fuzzy logic* en su esquema de clasificación, ha alcanzado una precisión del 96,49 % [29] en la tarea de clasificación de MNIST.

Por tanto, este trabajo sienta una base sólida para futuras investigaciones orientadas a aplicar modelos ART más avanzados en dominios más complejos, combinando su eficiencia estructural con técnicas modernas de representación de datos y selección de atributos. Esta exploración abre oportunidades prometedoras para aprovechar al máximo el aprendizaje dinámico de estas redes biológicamente inspiradas, que ofrecen una posible solución a los desafíos actuales del aprendizaje computacional.

BIBLIOGRAFÍA

- [1] Martin T. Hagan, Howard B. Demuth, Mark H. Beale, and Orlando De Jesús. *Neural Network Design*. Martin Hagan, 2014.
- [2] Curious.com. Best of 2015 #1: The science of "the dress", 2015. Consultado el 3 de mayo de 2025.
- [3] Stephen Grossberg. On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. *Biol. Cybernetics* 21, pages 145–159, 1976.
- [4] Stephen Grossberg. Adaptive pattern classification and universal recoding: I. parallel development and coding of neural feature detectors. *Biol. Cybernetics* 23, pages 121–134, 1976.
- [5] Stephen Grossberg. Adaptive pattern classification and universal recoding: II. feedback, expectation, olfaction, musions. *Biol. Cybernetics* 23, pages 187–202, 1976.
- [6] Stephen Grossberg. Neural facades: Visual representations of static and moving form-and-color-and-depth. *Mind & Language* 5, page N. 4, 1990.
- [7] James M. Keller, Derong Liu, and David B. Fogel. *Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation*. IEEE Press/Wiley, 2016.
- [8] Gail A. Carpenter and Stephen Grossberg. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, pages 54–115, 1987.
- [9] Yann LeCun and Corinna Cortes. MNIST handwritten digit database. <http://yann.lecun.com/exdb/mnist/>, 2010.

- [10] Eddy Torres-Constante, Julio Ibarra-Fiallo, and Monserrate Intriago-Pazmiño. *A New Approach for Optimal Selection of Features for Classification Based on Rough Sets, Evolution and Neural Networks*, pages 211–225. 08 2022.
- [11] Ilia Sucholutsky and Matthias Schonlau. Convar: Improving adaptive resonance theory for unsupervised image clustering. *Journal of Computational Vision and Imaging Systems*, 4:3, 12 2018.
- [12] Vladimir Rodríguez, José Francisco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, and Manuel S. Lazo-Cortés. Minreduct: A new algorithm for computing the shortest reducts. *Pattern Recognition Letters*, 136:46–52, 2020.
- [13] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [14] Yann LeCun and Corinna Cortes. MNIST handwritten digit database, 2010.
- [15] P. Y. Simard, D. Steinkraus, and J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of the International Conference on Document Analysis and Recognition (ICDAR)*, volume 3. Citeseer, 2003.
- [16] Stephen Grossberg. A path toward explainable ai and autonomous adaptive intelligence: Deep learning, adaptive resonance, and models of perception, emotion, and action. *Frontiers in Neurorobotics*, 14, 2020.
- [17] N. C. Kavuri and Madhusree Kundu. ART1 Network: Application in Wine Classification. *International Journal of Chemical Engineering and Applications*, 2(3), June 2011.
- [18] S. M. Soak, S. C. Chang, T. H. Shin, and B. H. Ahn. Music recognition system using ART-1 and GA. In *Proceedings of SPIE, Optical Pattern Recognition XIII*, volume 4734, pages 171–179. SPIE, 2002.
- [19] Teresa Serrano-Gotarredona and Bernabé Linares-Barranco. A Modified ART 1 Algorithm More Suitable for VLSI Implementations. *Neural Networks*, 1996. Submitted November 24, 1994; Accepted November 13, 1995.

- [20] Peng Li and Ma Xianxi. An Improved ART1 Neural Network Algorithm for Character Recognition. In *Proceedings of the IEEE International Conference*, pages 2946–2950. IEEE, 2010.
- [21] R. Thilepa and M. Thanikachalam. A New Process of Color Images Recognition by Using ART-1 Neural Network. *Computer and Information Science*, 4(1), January 2011.
- [22] V. D. Dmitrienko, A. Yu. Zakovorotnyi, S. Yu. Leonov, and I. P. Khavina. Neural Networks ART: Solving Problems with Multiple Solutions and New Teaching Algorithm. *The Open Neurology Journal*, 8:15–21, 2014.
- [23] Taisir Mohammad Eldos and Abdulaziz Suleiman Almazayd. Adaptive Resonance Theory Training Parameters: Pretty Good Sets. *Journal of Computer Science*, 6(12):1443–1449, 2010.
- [24] Ei Ei Khin, Amarottam Shrestha, and R. Sadananda. ART1: Similarity Measures. *Neural Processing Letters*, 6:109–117, 1997.
- [25] Gail A. Carpenter and Stephen Grossberg. Art 2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, 26(23), December 1987.
- [26] Hang Joon Kim, Jong Wha Jung, and Sang Kyoong Kim. On-line Chinese Character Recognition Using ART-based Stroke Classification. *Pattern Recognition Letters*, 17:1311–1322, 1996.
- [27] Gail A. Carpenter, Stephen Grossberg, and John H. Reynolds. Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4(5):565–588, 1991.
- [28] Gregory P. Amis and Gail A. Carpenter. Self-supervised ARTMAP. *Neural Networks*, 23:265–282, 2010.
- [29] Alejandro Baldominos, Yago Saez, and Pedro Isasi. A survey of handwritten character recognition with mnist and emnist. *Applied Sciences*, 9(15):3169, 2019.