



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

CENTRO UNIVERSITARIO UAEM VALLE DE MÉXICO

**Aprendizaje Incremental con Redes Neuronales
Artificiales**

TESIS

Que para obtener el Título de

INGENIERO EN SISTEMAS Y COMUNICACIONES

P r e s e n t a

**C. Fragoso García Sandra
C. González Hernández Luis Ángel**

Asesor: Dr. Víctor Manuel Landassuri Moreno

Atizapán de Zaragoza, Edo. de Méx. octubre 2022



Centro Universitario
UAEM Valle de México

Contents

1	Introduccion	4
2	Planeamiento	4
3	Objetivos	4
3.1	Objetivos Particulares	5
4	Justificación	5
5	Delimitación	6
6	Consecuencias	6
7	Marco Teorico	7
7.1	Revisión de la literatura	7
7.2	Aprendizaje Humano	7
7.2.1	Aprendizaje con Compresión	7
7.2.2	Aprendizaje Activo	8
7.3	Aprendizaje Incremental	8
7.3.1	Algoritmos de Aprendizaje Incremental	8
7.4	Redes Neuronales Artificiales	9
7.4.1	Redes Neuronales de Perceptrón Multicapa	9
7.4.2	Redes Neuronales Convolucionales	10
8	Hipotesis	11
9	Metodología	11
10	Cronograma de Actividades	12
11	Organización del Capitulado	12

List of Figures

1	Red Neuronal de Perceptrón Multicapa [8]	10
2	Esquema de una Red Convolutiva [3]	11
3	Diagrama de Gantt	12

1 Introduccion

El aprendizaje incremental es un método el cual a sido implementado en el área de la inteligencia artificial, ya que al realizar tareas específicas de dicha rama nos ayuda a optimizarla para que el algoritmo sea más eficiente. Cualquier tipo de aprendizaje puede ser considerado aprendizaje incremental si el problema a resolver tiene un entrenamiento simple, además este tipo de algoritmo es conocido como *algoritmo lineal sin memoria*, en la mayoría de los casos este tipo de aprendizaje es el preferido o favorito por los desarrolladores [5].

Las redes neuronales artificiales (RNAs) son procesos matemáticos los cuales son utilizados en el área de Machine Learning para la resolución de problemas no lineales, estos deben de pasar por una función de activación, la cual es una multiplicación entre los valores otorgados, al ser procesados por las capas que contenga la neurona, se obtendrá un valor distinto al de entrada.

Además son una distribución muy conocida de parte del Machine Learning, de otra manera, es el poder que tienen las computadoras para una buena estructura distribuida en paralelo y una buena habilidad de aprendizaje, este modelo computacional se define por medio de las neuronas biológicas las que son encargadas de que el ser humano pueda aprender o distinguir de distintos aspectos, este tipo de método es motivado para poder obtener la meta de un buen aprendizaje de máquina [8].

Un factor importante para esta rama es la pérdida de memoria, este es un problema biológico, el cual tanto afecta a los humanos como a las máquinas, es por eso que se han elaborado distintos experimentos para poder combatir esta problemática. Uno de estos es el caso de John Bullinaria, quien maneja la arquitectura de doble peso, ya que en su experimento da a notar que mejora la utilización del aprendizaje incremental, esto se logró con sistemas existentes como lo es Learn++.

Cabe mencionar que el experimento realizado fue un problema de generalización más comunes, pero se necesitan más evidencias de que usando esta metodología sirve para utilizarlo no solamente en tareas generalizadas, además se espera un mejor rendimiento [2].

2 Planeamiento

Indicar de forma general como es el funcionamiento de las redes neuronales en el uso general, como es que se produce la falta de memoria en estos modelos de predicciones. Poner en práctica los modelos de experimentación de John Bullinaria para comprobarlos con problemáticas más robustas y optimizarlo para que sea un método eficiente.

3 Objetivos

Diseñar una red neuronal para aprendizaje incremental basada en el principio de la memoria a corto y largo plazo, buscando usar más de dos categorías para el reconocimiento de dígitos.

3.1 Objetivos Particulares

1. Implementar el algoritmo de John A. Bullinaria para el reconocimiento de dígitos con aprendizaje a corto y largo plazo con los parámetros que él indica.
2. Obtener el conjunto de datos de Optical Digits y analizar que esté conforme al artículo de John A. Bullinaria.
3. Separar el conjunto de entrenamiento y de prueba de acuerdo a lo que se explica en el artículo base y hacer los experimentos para que se obtengan los mismo resultados.
4. Programar y modificar el algoritmo base, hacer que sean más de dos nodos, y repartir las tazas de aprendizaje proporcionalmente.
5. Probar la nueva implementación con el mismo conjunto de datos y ver si hay una diferencia significativa.

4 Justificación

Las redes neuronales apoyan a la resolución de distintos problemas, pero el Maching Learning tiene una deficiencia que es al momento de aumentar los datos a dichos modelos, la deficiencia que se obtiene es enorme que causa que los proyectos sean obsoletos [2]. Los resultados que se han obtenido no funcionan a la perfección, la memoria a corto plazo olvida poco pero va olvidando, y lo ideal sería que no olvidara, biológicamente nosotros no podemos hacer muchas modificaciones por lo mismo que implica, pero computacionalmente nada puede impedir que se pruebe con más configuraciones y llegar al punto en donde toda la información que llega se acumule y si no hay problema de almacenamiento que se siga acumulando y que no olvide, eso podría ser bueno en algunas situaciones.

Hacer este trabajo puede hacer que funcionen mejor las técnicas, pues ahorraría más energía en lugar de hacer entrenamientos muy grandes cada determinado tiempo e incluir todos los datos pasados de forma paulatina. Así como el ahorro de tiempo, de procesamiento, los tiempos de entrenamientos se reducirían y se reduciría la pérdida de información.

Existe una gran variedad de herramientas las cuales nos permiten codificar una red neuronal, existen 2 empresas importantes las cuales nos prestan sus servicios, las cuales son:

1. Microsoft.
2. Google.

Por el lado de Microsoft tenemos lo que es la plataforma de Azure que nos renta una maquina virtual donde podemos programar en Python.

Del lado de Google tenemos lo que es Google Colab que igual nos brinda una maquina virtual para realizar experimentos de Maching Learning, la única diferencia a Azure es que es gratuito, aquí también se puede programar en Python.

Como se observa en los dos se puede programar en Pyhton y es porque este lenguaje es una herramienta de software libre que no requiere licencia, es relativamente fácil poder depurar un código y permite acelerar más el desarrollo de

aplicaciones, a diferencia de otros lenguajes más estructurados como c o java, además tiene más librerías para el desarrollo de Maching Learning.

TensorFlow es una librería de Python que permite construir y entrenar redes neuronales para detectar patrones y razonamientos usados por los humanos.

Keras es un framework de alto nivel para el aprendizaje, escrito en Python y capaz de correr sobre los frameworks TensorFlow. Fue desarrollado con el objeto de facilitar un proceso de experimentación rápida. Diseñado para construir por bloques la arquitectura de cada red neuronal, incluyendo redes convolucionales y modelos recurrentes, que son las que permiten, junto a los bloques "más tradicionales", entrenar aprendizaje profundo.

5 Delimitación

En la siguiente investigación solamente se van a utilizar redes neuronales artificiales, cabe mencionar que este no es el único tipo de red, porque también se tiene lo que es [7]:

- Redes Neuronales Monocapa.
- Redes Neuronales Perceptrón Multicapa (MLP).
- Redes Neuronales Convulcionales (CNN).
- Redes Neuronales Recurrentes (RNN).
- Redes de Base Radial (RBF).

Pero para este experimento se va a utilizar ANN que es el algoritmo que por el momento nos beneficiará, cabe mencionar que no se usarán algoritmos genéticos, ya que si se implementa, se estará optimizando y el objetivo principal es utilizar el aprendizaje incrementado para que acepte más datos de entrenamiento.

6 Consecuencias

Si el experimento funciona a la perfección ocurrirá lo siguiente:

1. Habrá menos olvido.
2. Los procesos tardaran menos tiempo.

Esto sucederá porque se van a poder ingresar más datos a nuestro almacén sin tener que volver a codificar nuestro modelo.

Los modelos de predicción van a ser más precisos, porque se podrán ingresar datos mensualmente o hasta semanalmente, esto provocará que el proyecto esté trabajando con datos actuales.

Pueden funcionar para proyectos y predicciones tan simples tanto la preidcción climatológica hasta predicciones de la bolsa de valores y predicción del Bitcoin.

7 Marco Teórico

7.1 Revisión de la literatura

El humano tiene una forma de aprendizaje muy particular, la cual se basa del estudio, donde lee, escribe y practica acerca de su tema de interés, pero dicho aprendizaje se puede ir olvidando, esto es una acción muy común que a cualquier persona le sucede. Existen estudios donde se comenta que existen tres motivos del porque se olvidan las cosas, proviene parte de la regularización de las emociones, el como se adquirieron los conocimientos, y porque el olvido es un proceso por el cual el ser humano transita a lo largo de su vida [9]. Pero cabe mencionar que esto no es lo único que causa la pérdida de memoria, ya que existe la déficit de memoria.

7.2 Aprendizaje Humano

Al momento de hablar del aprendizaje humano, se debe de hablar de la ciencia cognitiva, que es quien se encarga de descubrir esta incógnita, esta ciencia lo estudia de un modo multidisciplinario, el cual abarca las áreas de [1]:

- La antropología.
- La lingüística.
- La filosofía.
- La sicología del desarrollo.
- La ciencia de la computación.
- La neurociencia.

Con el método de esta ciencia se pueden descubrir dos tipos de aprendizaje que son:

1. Aprendizaje con Compresión.
2. Aprendizaje Activo.

7.2.1 Aprendizaje con Compresión

La comprensión es una actividad la cual se ha generado al momento de realizar cualquier tipo de lectura.

Al hablar de este tema nos enfocamos en el ámbito estudiantil que es donde más se maneja esta táctica, esto es una practica algo compleja, sistemática y organizada, ya que nos da el significado de la literatura, gracias a esto se puede obtener el contexto de la literatura.

Al conocer esto podemos decir con seguridad que para cualquier tipo de aprendizaje la comprensión es una parte primordial [10].

7.2.2 Aprendizaje Activo

El aprendizaje de la forma en la que se conocen no es del todo efectiva, se dice esto porque parte del sistema de educación porque la manera correcta es con el principio de *belongingness*, el cual esta asociado al estímulo con su respuesta, y esto es lo más importante para que el ser humano pueda aprender cualquier cosa.

Este tipo de aprendizaje se basa en la recepción de conocimientos y la práctica donde se ponen en marcha los conocimientos adquiridos.

Otro concepto importante aquí es la tautología doble (*selbsttätiges Lernen*), que en palabras informales es convertirse en autodidacta, se puede observar que esto pertenece a dicho aprendizaje, porque usa el principio mencionado anteriormente [6].

7.3 Aprendizaje Incremental

Con el pasar de los años la tecnología a evolucionado, eso quiere decir que el Aprendizaje Automático se ha actualizado, que la cantidad de datos va aumentando con más frecuencia.

Se puede verificar como *"Una tarea de aprendizaje es incremental si los ejemplos de entrenamiento usados para resolverla están disponibles en horas extras, generalmente uno a la vez"* [5], si los resultados no se necesitan de manera urgente, este tipo de trabajos serán resueltos por algoritmos de aprendizaje no incremental.

Una área donde esto es de mucha utilidad es la *Robótica* porque este necesita estar en constante entrenamiento [5].

Dicha forma de aprender fue inspirada en la forma que el humano aprende y esta es una forma más rápida, fue por esto que fue adoptada por el aprendizaje maquina.

Con el paso del tiempo se ha convertido en un paradigma del aprendizaje automático, aquí el aprendizaje toma el lugar de nuevos ejemplos para juntarlos y conforme van aprendiendo estos toman el lugar de los ejemplos ya aprendidos [8].

7.3.1 Algoritmos de Aprendizaje Incremental

"Un algoritmo de aprendizaje es incremental si, para cualquier muestra de entrenamiento dada:

$$e_1, \dots, e_s \quad (1)$$

, produce un secuencia de hipótesis

$$h_0, h_1, \dots, h_n \quad (2)$$

, tal que h_{i+1} depende solo de h_i y del ejemplo actual e " [5], como se observa, estos son algoritmos que permiten a la inteligencia artificial poder realizar actividades de predicción de una manera más eficaz.

Un ejemplo del uso de esta rama es el proyecto *COBWEB*, donde se trata de categorizar el número de Clúster y la pertenencia de dichas categorías por medio de una métrica probabilística global, esto lo realiza por medio de que se agrega una nueva categoría, este proceso lo que realizará es actualizar todas las probabilísticas con los nuevos datos recabados [4].

7.4 Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) son procesos los cuales contienen simples unidades de procesamiento.

Como podemos darnos cuenta, al mencionar RNA lo primero que se nos viene a la mente es el procesamiento biológico por el que transita el cerebro humano. El método principal de las redes neuronales es sacarle el máximo poder a los algoritmos de aprendizaje máquina, ya que las redes neuronales tienen un antecedente biológico.

Tiene demasiadas utilidades las cuales ayudan a distintas problemáticas, algunas de las funciones que tiene son [8]:

1. No linealidad.
2. Mapeo entrada-salida.
3. Aprendizaje robusto a errores en los datos de entrenamiento.
4. Entre otros.

Existen varios tipos de Redes Neuronales tales como:

- Redes Neuronales de Perceptrón Multicapa.
- Redes Neuronales Convolucionales.
- Entre otras.

7.4.1 Redes Neuronales de Perceptrón Multicapa

Al momento de mapear el progreso de una red nos sirve para conocer buscar un camino adecuado entre la capa de entrada y salida, esto se puede hacer con problemas de una sola neurona, pero no se pueden resolver problemas no lineales, es aquí donde entran los perceptrones multicapas, los cuales rompen con esta limitación.

Estas neuronas están compuestas de 3 capas:

1. Capas de entrada.
2. Capas ocultas.
3. Capas de salida.

Los datos de entrada se van a ir propagando capa por capa, hasta llegar a la capa final, donde ya pasó por una función de activación donde se rompió la no linealidad, para poder llegar a un resultado [8].

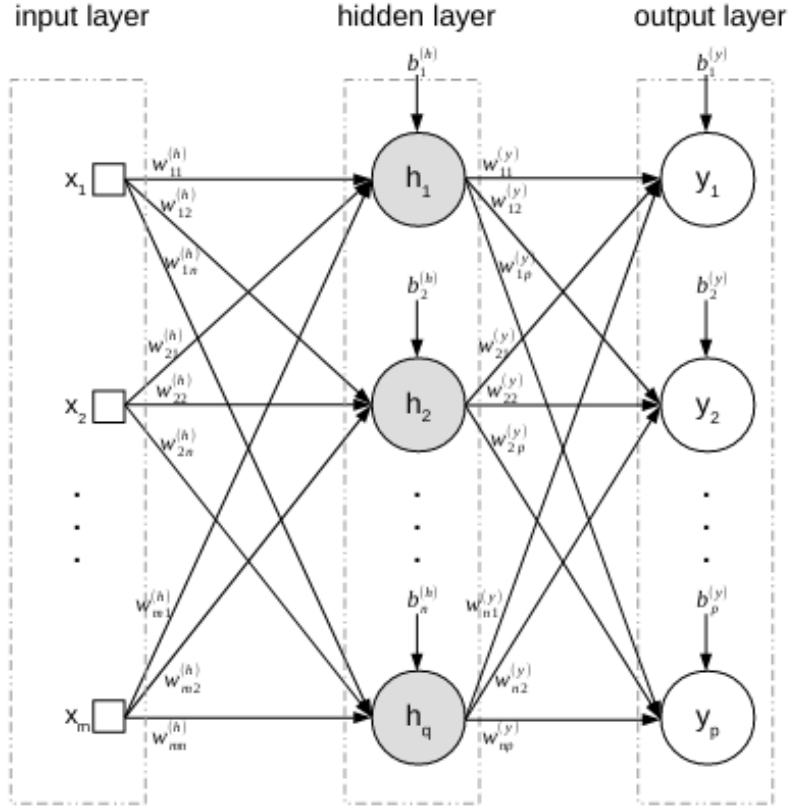


Figure 1: Red Neuronal de Perceptrón Multicapa [8]

7.4.2 Redes Neuronales Convolucionales

Este tipo de red trabaja con el uso de imágenes, por lo general de alta calidad, el único problema que se tiene al momento de que sean de alta resoluciones son:

- El tiempo de entrenamiento sea enorme.
- El tiempo de testeo sea muy tardío.

Consta de diversas multicapas alternadas, al final tiene una red perceptrón multicapa. La entrada de una red convolucional, con diferentes medidas en altura y anchura de imagen, para el uso de los proyectos se trabajan en escalas de grises, las cuales contienen filtros y cada filtro tiene distintos rasgos y características de tamaño. Cada capa es submuestreo de mínimo a máximo, muestra donde se toman valores desde 2 imágenes pequeñas hasta no mas de 5 imágenes grandes.

Antes o después del submuestreo se aplica la activación sigmoideal para cada mapeo de rasgos [3].

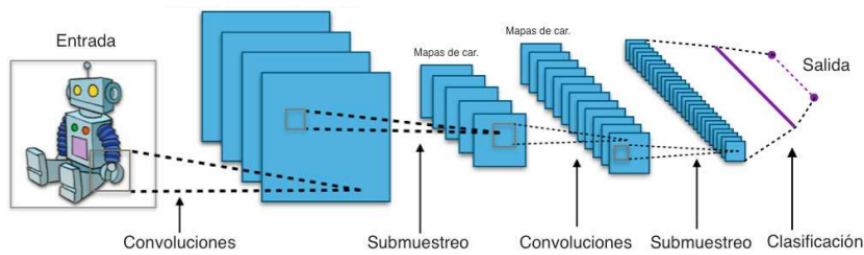


Figure 2: Esquema de una Red Convolutional [3]

8 Hipotesis

Al tener más conexiones en un modelo de red neuronal usualmente se tendrá mejor ajuste de los problemas, ya que al momento de predecir este será mucho más eficiente de manejar.

9 Metodología

El primer paso a realizar en esta investigación es recrear el código de John A. Bullinaria, para esto se va a utilizar el repositorio de Optical Digits, al momento de obtener estos datos ya se va a poder comenzar con la programación en Python.

Al momento de comprobar que el código funciona como el algoritmo de John, se realizarán modificaciones que permitirán repartir las tazas de aprendizaje con estos valores.

Para obtenerlo se usarán metodologías como el Backpropagation y funciones de activación tal como la Sigmoidal.

Cuando los dos proyectos se tengan, se realizará una comparación, donde se vera cual de estos dos experimentos es más eficaz en proyectos de la vida real.

10 Cronograma de Actividades

Algoritmo de John Bullinaria

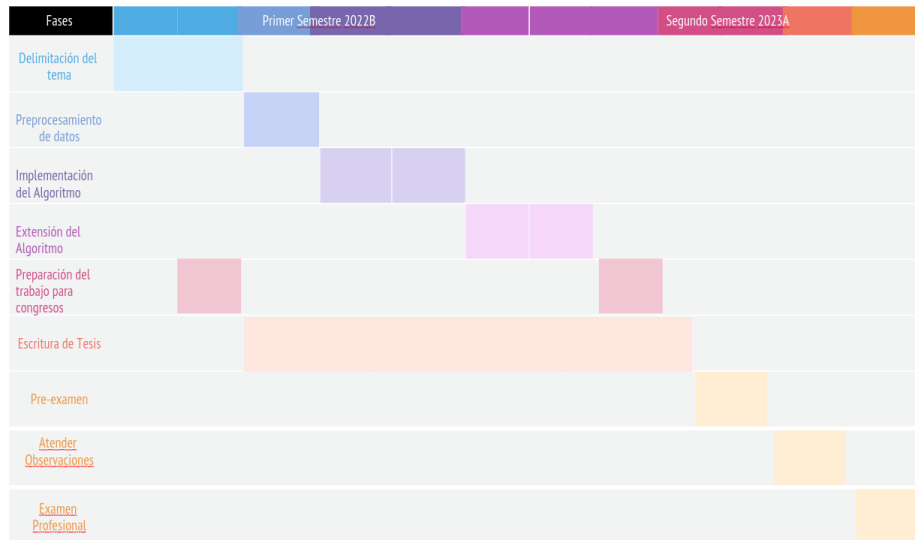


Figure 3: Diagrama de Gantt

11 Organización del Capitulo

En el capítulo 2 se verá lo que es el aprendizaje humano y el aprendizaje incremental con sus algoritmos, se describirán las redes neuronales artificiales.

En el capítulo 3 se implementará el artículo de John A. Bullinaria, como funciona, resultado que da al pasar los datos que dice para comprobar que funciona como menciona en su artículo. En el capítulo 4 se explicará como se hará la modificación a su algoritmo, cuantas capas se van a poner, como se van a repartir las tazas de aprendizaje.

Posteriormente en el capítulo 5 se mostrará una comparación de los resultados de ambos trabajos. En el capítulo 6 se verán las conclusiones y trabajo futuro.

References

- [1] J Bransford, A Brown, and R Cocking. “Cómo aprende la gente: cerebro, mente, experiencia, y escuela.” In: *Revista del Instituto de Matemática y Física* (2000), pp. 44–64.
- [2] John Bullinaria. “Evolved Dual Weight Neural Architectures to Facilitate Incremental Learning.” English. In: *Proceedings of the International Joint Conference on Computational Intelligence (IJCCI 2009)* ; Conference date: 01-11-2009. Nov. 2009, pp. 427–434.
- [3] Jaime Durán Suárez. “Redes neuronales convolucionales en R: Reconocimiento de caracteres escritos a mano.” In: (2017).

- [4] Douglas H Fisher. “Knowledge acquisition via incremental conceptual clustering.” In: *Machine learning* 2.2 (1987), pp. 139–172.
- [5] Christophe G. Giraud-Carrier. “A Note on the Utility of Incremental Learning.” In: *AI Commun.* 13 (2000), pp. 215–224.
- [6] Günter L. Huber. “Aprendizaje activo y metodologías educativas.” In: (2008). URL: http://www.revistaeducacion.mec.es/re2008/re2008_04.pdf.
- [7] family=Royo given i=P. given=Paloma. *Qué son las redes neuronales y cuál es su aplicación en el marketing*. Sept. 2021. URL: <https://artyco.com/que-son-las-redes-neuronales-y-cual-es-su-aplicacion-en-el-marketing/>.
- [8] Yuan Liu. “Incremental learning in deep neural networks.” MA thesis. 2015.
- [9] Simon Nørby. “Why Forget? On the Adaptive Value of Memory Loss.” In: *Perspectives on Psychological Science* 10 (2015), pp. 551–578.
- [10] Karel Pérez Ariza and José Emilio Hernández Sánchez. “Aprendizaje y comprensión. Una mirada desde las humanidades.” In: *Humanidades Médicas* 14.3 (2014), pp. 699–709.