

Aprendizaje Incremental para la Tarea de Reconocimiento de Dígitos con Redes Neuronales Artificiales

by

González Hernández Luis Ángel

CUUAEM VM
UAEM Diciembre 2022

Resumen

El aprendizaje incremental es un área de la Inteligencia Artificial la cual permite agregar nuevo conocimiento a un modelo (e.g. Redes Neuronales Artificiales) sin la necesidad de entrenar el modelo con toda la información histórica de la tarea en cuestión [1]. En el presente trabajo de investigación se ocupará el modelo de Redes Neuronales Artificiales enfocada en la clasificación de dígitos escritos a mano usando el algoritmo de entrenamiento de backpropagation, con redes Multi capa Perceptron y duplicación de pesos múltiples simulando memoria a corto y largo plazo para mejorar los resultados presentados en [1].

Dedicatoria

Agradecimientos

Sus agradecimientos

Índice general

1. Introducción	1
2. Revisión de la Literatura	4
2.1. Redes Neuronales Artificiales	4
2.1.1. Función de Activación	5
3. Metodología	7

Índice de figuras

2.1. Red Neuronal Artificial Básica	4
2.2. Función Escalonada	6

Índice de tablas

Lista de Acrónimos

Capítulo 1

Introducción

La inteligencia artificial (IA) es una área del conocimiento que se enfoca en poder hacer máquinas que se enfocan en comportamiento y razonamiento humano, para que en momento dado, se pueda interactuar con una máquina. Así mismo, también es posible pensar que mucho del desarrollo en el área de inteligencia artificial, es el poder tener mejores herramientas que ayuden a las actividades diarias.

En este sentido, un área de la IA es el llamado Aprendizaje Máquina, donde se estudian algoritmos que permiten aprender de forma automática una tarea. Así, una de las técnicas más conocidas en la actualidad, dentro del área de IA son las Redes Neuronales Artificiales (RNAs), siendo técnicas que realizan procesos matemáticos para poder aprenderse tareas a resolver. Algunas áreas en las que son útiles las RNAs son en el aprendizaje de tareas no lineales, como la predicción de la capacidad de la red 5G, basada en el tráfico diario de este [2] o clasificación, por ejemplo la clasificación de metales y rocas por medio de RNAs y lógica difusa [3].

Las RNAs están formadas por neuronas artificiales que simulan a las biológicas. Así los procesos químicos que suceden en el cerebro, se simulan computacionalmente a través de señales que viajen a través de las neuronas artificiales, de aquí en adelante simplemente se referirá a ellas como "neuronas". Las neuronas en una RNA cuentan con una

estructura distribuida en paralelo, presentando una buena habilidad de aprendizaje [4].

Dentro del aprendizaje máquina cuando una técnica, por ejemplo RNA, se enfoca en aprender una tarea, se conoce como *algoritmo lineal sin memoria*, siendo uno de los métodos más empleados desde el inicio de las RNAs [5]. Sin embargo, si es necesario incorporar nueva información del problema, es necesario volver a entrenar todo el modelo, considerando toda la información existente, esto es, la anterior y la nueva que acaba de llegar, es ahí donde nace el concepto de Aprendizaje Incremental, siendo un área enfocada en poder incorporar información del problema en cuestión, sin tener que volver a re-entrenar todo el modelo.

Derivado del aprendizaje incremental se desprende el concepto de memoria dentro de la IA, analizando como un algoritmo de aprendizaje máquina puede olvidar la información que se usó en un entrenamiento previo al entrenar con información más reciente. Si se hace la analogía con los humanos, la memoria es un factor importante para estudiar considerando la perdida de información aprendida, así, este es un problema biológico, el cual tanto afecta a los humanos como a las máquinas. Por ello, se han elaborado distintos experimentos para poder combatir esta problemática. Uno de estos es el caso de [1], el cual propone el manejo de RNAs con pesos dobles, donde la primer capa de pesos esta enfocada a comportarse como memoria a corto plazo, y la segunda como memoria a largo plazo. Los experimentos mostrados en [1] permiten notar un mejora en tareas de aprendizaje incremental, teniendo menos pérdida de información en comparación de implementaciones anteriores como el algoritmo Learn++ [6, 7].

anteriores como el algoritmo Learn++ [6, 7].

Así, el presente trabajo de investigación esta enfocado en poder explorar nuevas configuraciones. Así, el presente trabajo de investigación esta enfocado en poder explorar nuevas configuraciones de pesos duplicados para poder extender el trabajo previamente presentado en [1]. de pesos duplicados para poder extender el trabajo previamente

presentado en [1].

Capítulo 2

Revisión de la Literatura

2.1. Redes Neuronales Artificiales

Las redes neuronales artificiales (RNA) son modelos computacionales de la Inteligencia Artificial los cuales contienen simples unidades de procesamiento llamadas neuronas. Ellas se inspiran en el cerebro humano, tomando como base la conectividad entre neuronas y el aprendizaje que pueden tener. Un perceptron o neurona (artificial) solamente resuelve problemas lineales y tiene la siguiente forma:

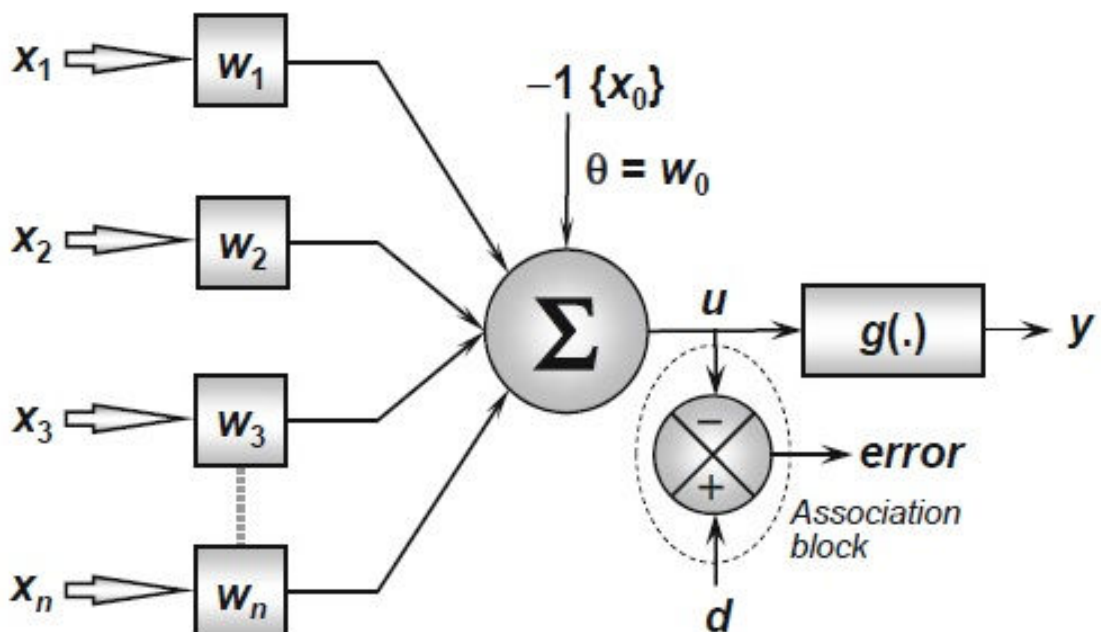


Figura 2.1: Red Neuronal Artificial Básica

Donde Σ es la representación matemática de la neurona., x_1, x_2, \dots, x_n son las variables de entrada a la red. w_1, w_2, \dots, w_n son los pesos con los cuales se van a ponderar las entradas, es decir multiplicar cuando la información entra en la neurona. Posterior a multiplicar el peso por la entrada correspondiente, se suman todos esos valores $w_1x_1 + w_2x_2 + w_3x_3$.

Al revisar esta formula, se puede observar que se parece a la operación de una regresión la cual es: $y = w_0 + w_ix_i$, de esta forma, internamente la neurona realiza una regresión lineal. En su contraparte, el parámetro que permite a la neurona trazar una recta cruzando el eje y en el plano cartesiano (eje de las ordenadas), a ello se conoce como sesgo (del inglés *bias*), este valor se agrega a la conexión, el cual usualmente se le da un valor de 1. Agregando este nuevo valor a la fórmula, queda de la siguiente manera: $y = \Sigma w_ix_i + w_0b$, donde b es el sesgo.

Un inconveniente del uso de una sola neurona para experimentos es que solo va a resolver ejercicios parecidos a la puerta lógica AND u OR.

Existen problemáticas de solo usar una sola neurona e.g problemas de tipo compuerta XOR.

Para solucionarlo se usan dos o más neuronas, además de la función de activación, que es la que permite pasar la información de una neurona a otra, en un rango especificado, y la cual se describirá en la siguiente sección.

2.1.1. Función de Activación

Dicho método se utiliza cuando el modelo de RNA contiene dos o más neuronas. Esta función lo que provoca es dar al modelo una salida no lineal, para eso la segunda fórmula presentada es distorsionada para quedar de la siguiente manera: $f(w_1x_1 + w_2x_2 + w_3x_3 + b_0)$ para el caso de 3 entradas.

Al hablar de funciones de activación se deben de comentar las más comunes, como lo es la función escalonada.

Tenemos la función escalonada, la cual se representa con la siguiente formula:

$$f(x) = \begin{cases} 0 & : x < 0 \\ 1 & : x \geq 0 \end{cases}$$

Su representación grafica queda de la siguiente manera:

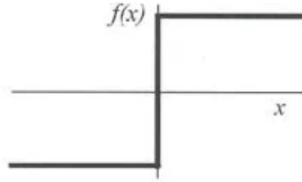


Figura 2.2: Función Escalonada

Capítulo 3

Metodología

El primer paso a realizar, es recrear el código mostrado en [1], el describe la implementación de una red neuronal multicapa usando el algoritmo de entrenamiento back-propagation en el lenguaje de programación Python. Así mismo, se utilizará el conjunto de datos de Optical Digits, en donde se tendrá que preprocesar los datos, para eliminar registros inválidos.

Posteriormente, se implementará una extensión del código, donde se experimentará con mas de dos capas de pesos duplicados para mejorar la tasa de olvido de información al momento de usar el aprendizaje incremental. Para ello se explorará incrementando gradualmente el número de capas de pesos duplicados.

Finalmente, cuando los resultados se obtengan se realizará una comparación, de los resultado del algoritmo base con los resultados del algoritmo extendido.

Bibliografía

- [1] J. A. Bullinaria, “Evolved dual weight neural architectures to facilitate incremental learning,” in *IJCCI*, pp. 427–434, 2009.
- [2] B. Zhao, T. Wu, F. Fang, L. Wang, W. Ren, X. Yang, Z. Ruan, and X. Kou, “Prediction method of 5g high-load cellular based on bp neural network,” in *2022 8th International Conference on Mechatronics and Robotics Engineering (ICMRE)*, pp. 148–151, IEEE, 2022.
- [3] L. A. C. Salazar, D. J. M. Aldana, and J. A. C. Montes, “Implementación de redes neuronales y lógica difusa para la clasificación de patrones obtenidos por un sónar,” in *2013 II International Congress of Engineering Mechatronics and Automation (CIIMA)*, pp. 1–6, IEEE, 2013.
- [4] Y. Liu, “Yuan liu incremental learning in deep neural networks,” 2015.
- [5] C. G. Giraud-Carrier, “A note on the utility of incremental learning,” *AI Commun.*, vol. 13, pp. 215–224, 2000.
- [6] A.-J. Li, “An improved algorithm for incremental learning learn++,” in *2008 International Conference on Wavelet Analysis and Pattern Recognition*, vol. 1, pp. 310–315, IEEE, 2008.
- [7] R. Elwell and R. Polikar, “Incremental learning of concept drift in nonstationary environments,” *IEEE Transactions on Neural Networks*, vol. 22, no. 10, pp. 1517–1531, 2011.