

UNIVERSIDAD DE ORIENTE

NÚCLEO DE SUCRE — EXT. CARÚPANO

ESCUELA DE CIENCIAS

DEPARTAMENTO DE INFORMÁTICA

CÁTEDRA DE PROGRAMACIÓN EMERGENTE I

**ANÁLISIS DE LA BASE DE DATOS IMDB**

**A TRAVÉS DE LA APLICACIÓN DE**

**MAPAS AUTO ORGANIZATIVOS**

PROFESOR: BACHILLER:

RENAN SALAZAR JESÚS ORDOSGOITTY

C.I: 27 572 434

CARÚPANO, JUNIO DE 2019

ÍNDICE

[INTRODUCCIÓN 3](#_Toc12418498)

[1 PLANTEAMIENTO DEL PROBLEMA 4](#_Toc12418499)

[2 BASES TEÓRICAS 4](#_Toc12418500)

[Red Neuronal Artificial 4](#_Toc12418501)

[Mapa auto organizado de características 4](#_Toc12418502)

[Aprendizaje 5](#_Toc12418503)

[Regla de aprendizaje 5](#_Toc12418504)

[Aprendizaje competitivo 6](#_Toc12418505)

[Aprendizaje no supervisado 7](#_Toc12418506)

[3 SOFTWARE EMPLEADO 7](#_Toc12418507)

[4 PROCEDIMIENTOS 7](#_Toc12418508)

[5 CONCLUSIONES Y RECOMENDACIONES 8](#_Toc12418509)

[REFERENCIAS 9](#_Toc12418510)

# INTRODUCCIÓN

# 1 PLANTEAMIENTO DEL PROBLEMA

# 2 BASES TEÓRICAS

Las bases teóricas son el grupo de argumentos que fundamentan la formulación y el desarrollo de una investigación, en este sentido, implican una conceptualización amplia de las teorías que basan el punto de vista o enfoque adoptado, con el fin de sustentar al presente proyecto [1]. A continuación, se hace mención a las teorías y conceptos relacionados al tema de investigación.

### Red Neuronal Artificial

Para obtener una definición de red neuronal [2] tenemos que hacer uso del concepto matemático de grafo. A través de este término, podemos definir una red neuronal de la siguiente forma como un grafo dirigido con las siguientes propiedades:

1. A cada nodo se le asocia un variable de estado .
2. A cada conexión de los nodos y se le asocia un peso .
3. A cada nodo se le asocia un umbral .
4. Para cada nodo i se define una función , que depende de los pesos de sus conexiones, del umbral y de los estados de los nodos a él conectados. Esta función proporciona el nuevo estado del nodo.

### Mapa auto organizado de características

El mapa de Kohonen, SOM (*self-organizing map,* mapa auto organizativo) o SOFM (*self-organizing feature map,* mapa auto organizado de características) es un tipo de red neuronal artificial que es entrenada a través del aprendizaje no supervisado para producir una representación discretizada, de baja dimensión (típicamente, dos dimensiones) del espacio de entrada de un conjunto de muestras de entrenamiento, el cual se denomina mapa [5]; es un método para la reducción de la dimensionalidad de la data. Los mapas auto organizativos difieren de otros tipos de RNAs ya que aplican métodos de aprendizaje competitivo en contra de procedimientos basados en la corrección de errores (tales como la propagación hacia atrás con la gradiente descendiente), y en ese sentido, usan una función de vecindad para preservar las propiedades topológicas del espacio de entrada.

La parte visible de un mapa auto organizado es el mapa espacial, el cual consiste de *n* componentes llamados nodos o neuronas. El mapa espacial se encuentra definido desde el inicio, generalmente como una región finita de dos dimensiones donde los nodos son colocados en una malla configuración hexagonal o rectangular. Cada nodo está asociado a un vector de pesos, el cual es una posición en el espacio de entrada; esto quiere decir, tiene la misma dimensión de cada vector de entrada [5]. Mientras que los nodos en el mapa espacial se encuentran fijos, el entrenamiento consiste en mover los vectores pesados hacia la data de entrada (reduciendo una distancia métrica) sin modificar la topología desde el mapa espacial. Después de todo, el mapa auto organizativo describe un mapeo desde una dimensional más alta de un espacio de entrada a un mapa espacial de baja dimensión. Una vez entrenado, el mapa puede clasificar un vector del espacio de entrada encontrando vector de pesos más cercano (menor distancia métrica) al espacio de entrada vectorial.

El mapa auto organizado de características fue diseñado de manera que fuese una alternativa viable a las arquitecturas de redes neuronales más tradicionales. Es posible preguntar qué tan “neuronal” es un mapa ya que su descripción analítica ha sido más desarrollada en el ámbito tecnológico más que en lo biológico [7]. Pero los resultados alcanzados tras el aprendizaje parecen muy naturales, indicando que el proceso adaptativo que trabaja en el mapa es similar a aquellos encontrados en el cerebro.

### Aprendizaje

Puede definirse como el proceso por el que se produce el ajuste de los parámetros libres de la red a partir de un proceso de estimulación por el entorno que rodea a la red [4]. En la mayoría de los casos el aprendizaje consiste simplemente en determinar un conjunto de pesos sinápticos que permita a la red realizar correctamente el tipo de procesamiento deseado.

Al construir un sistema neuronal, se parte de un cierto modelo de neurona y de una determinada arquitectura de red, estableciéndose los pesos sinápticos iniciales como nulos o aleatorios. Para que la red resulte operativa es necesario entrenarla. En el caso de los mapas auto organizados, el entrenamiento o aprendizaje se puede llevar a cabo simplemente a través del modelado de las sinapsis; que consiste en modificar los pesos sinápticos siguiendo la regla de aprendizaje, construida a partir de la optimización de una función de error o coste, que mide la eficacia actual de la operación de la red. En el caso de los mapas auto organizados[5], la fórmula para la actualización de una neurona con un vector de pesos viene siendo:

### Regla de aprendizaje

Es un algoritmo que puede ser empleado para cambiar y, en consecuencia, entrenar la red neuronal, de manera que la red genere una salida deseada para una entrada dada [3]. Con respecto a las redes de Kohonen, utilizamos un algoritmo diseñado para calcular la similitud entre neuronas y su modificación, el cual conoceremos a continuación.

### Aprendizaje competitivo

Es un tipo de algoritmo para el cual se entrena una sola neurona en cada época [4] [5] lo cual significa que la representación de cada zona del espacio de entrada está concentrada por neuronas, no distribuida (como suele suceder en otras redes neuronales tales como el perceptrón multicapa).

Generalmente, el algoritmo es descrito de la siguiente manera:

1. Hacer un mapa de neuronas con vectores de pesos aleatorios.
2. Tomar un vector de entradas e iterar por cada neurona del mapa.
   1. Calcular la distancia entre el vector de entrada y los vectores de pesos de las neuronas del mapa, es decir:
   2. Mantener la neurona que ha tenido la menor distancia; esta neurona será el *BMU*.
3. Actualizar los pesos de la neurona ganadora y sus vecinas empleando la primera función.
4. Incrementar y volver al paso 2, mientras .

Dónde:

* Es la iteración actual.
* Es la cantidad total de iteraciones.
* Es el índice del vector de entrada en el conjunto de datos de entrada .
* Es un vector de entrada de índice T del conjunto de datos de entrada.
* Es el índice de una neurona en el mapa.
* Es el vector de pesos de la neurona .
* Es el índice del BMU en el mapa.
* Es la función de vecindad.
* Es el ritmo de aprendizaje de acuerdo al progreso de las iteraciones.

### Aprendizaje no supervisado

Este tipo de aprendizaje [2] se puede describir genéricamente como la estimación de la función densidad de probabilidad que describe la distribución de patrones  (espacio de entrada).

En este aprendizaje se presentan a la red multitud de patrones sin adjuntar la respuesta que deseamos. La red, por medio de la regla de aprendizaje, estima , a partir de lo cual podemos reconocer regularidades en el conjunto de entradas, extraer rasgos o agrupar patrones según su similitud (clustering).

Cabe destacar que los métodos no supervisados se suelen usar en lo denominado *análisis de datos exploratorio* [4], es decir, en una fase del análisis de los datos, cuando no se sabe de antemano cuáles son los grupos naturales que se forman, y se quiere visualizar la abundancia y la relación que hay entre los grupos "naturales"; se puede decir que una de sus principales aplicaciones es la visualización de datos multidimensionales, porque un algoritmo no supervisado actúa como una proyección de un espacio multidimensional a otro de dimensiones visualizables. También se pueden usar como fase inicial de algoritmos de aprendizaje supervisados: un algoritmo como el *k-medias* o el mismo SOM se pueden usar para inicializar algoritmos de aprendizaje supervisado tales como el LVQ (*Learning Vector Quantization*).

# 3 SOFTWARE EMPLEADO

# 4 PROCEDIMIENTOS

# 5 CONCLUSIONES Y RECOMENDACIONES

# REFERENCIAS

[1] Fidias, A. (2012). *El Proyecto de Investigación.* (6a Ed.). Caracas: Episteme.

[2] *Conceptos básicos sobre Redes Neuronales*. Consultado el 9 de mayo de 2019. Disponible en: <http://grupo.us.es/gtocoma/pid/pid10/RedesNeuronales.html>

[3] Kriesel, D. (2013). *A Brief Introduction to Neural Network*. (2a Ed.). Bonn: Autor.

[4] J. J. Merelo. *Mapa autoorganizativo de Kohonen*. Consultado el 18 de mayo de 2019. Disponible en: <http://geneura.ugr.es/~jmerelo/tutoriales/bioinfo/Kohonen-2005.html>

[5] Wikipedia. *Self-organizing map*. Consultado el 18 de mayo de 2019. Disponible en: <https://en.wikipedia.org/w/index.php?title=Self-organizing_map&oldid=892805093>

[6] Garlan, D., Shawn, M. (1994). *An Introduction to Software Architecture*. Pittsburgh, PA: Carnegie Mellon University.

[7] Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of the IEEE, 78*, (9), 1464-1480

[8] Kokatjuhha, J. *How to build a data science project from scratch*. Consultado el 20 de junio de 2019. Disponible en: <https://www.freecodecamp.org/news/how-to-build-a-data-science-project-from-scratch-dc4f096a62a1/>