Universitat de les Illes Balears

Facultat de Psicologia

Memòria del Treball de Fi de Grau

El uso del Perceptrón Multicapa para la clasificación de patrones en conductas adictivas

Marta Vidal González

Grau de Psicologia

Any acadèmic 2014-15

DNI de l'alumne: 43227812W

Treball tutelat per Juan José Montaño Moreno
Departament de Psicologia

√ S'autoritza la Universitat a incloure el meu treball en el Repositori Institucional per a la seva consulta en accés obert i difusió en línia, amb finalitats exclusivament acadèmiques i d'investigació

Paraules clau del treball:
Redes Neuronales, Perceptrón Multicapa, Clasificación

Índice

Resumen …..... 4

Introducción …..... 5

Revisión histórica …..... 5

Concepto …..... 9

Fundamentos biológicos …..... 11

Elementos de una red neuronal 12

Mecanismos de aprendizaje …..... 14

Tipos de red neuronal …..... 14

Perceptrón Multicapa …..... 17

Aplicaciones de las Redes Neuronales Artificiales 19

Uso de las Redes Neuronales Artificiales en las Conductas Adictivas

…..... 21

Objetivo …..... 22

Metodología …..... 22

Matriz de datos …..... 22

Instrumentos …..... 25

Procedimiento …..... 25

Resultados …..... 29

Grupo de entrenamiento …..... 29

Grupo de validación …..... 31

Grupo test …..... 32

Conclusión …..... 35

Bibliografía …..... 36

ANEXO 1 …..…..... 45

ANEXO 2 …..…..... 46

Etapas de funcionamiento 48

Etapa de aprendizaje 48

Etapa de funcionamiento 50

Etapa test …..... 51

Consideraciones a tener en cuenta 51

ANEXO 3 …..... 54

Resumen

En este trabajo se quiere demostrar el funcionamiento del Perceptrón Multicapa como modelo clasificador de las conductas adictivas creando un modelo de red con el sistema *Sensitive Neuronal Network* y usando una muestra de 628 sujetos divididos en tres grupos: entrenamiento (314), validación (157) y test (157). Los resultado obtenidos son en el grupo test son que el 56,05% de los participantes recaen en las drogas y el 43,95% no recaen. La eficacia del modelo seleccionado es del 74%.

Palabras clave: Redes Neuronales Artificiales, historia, modelos, Perceptrón Multicapa, fundamentos.

Introducción

El cerebro humano es un sistema de cálculo muy complejo, puede llevar a cabo procesamientos que a primera vista parecen sencillos, como por ejemplo, el reconocimiento de imágenes. Esta capacidad que tiene el cerebro humano para pensar, recordar y resolver problemas ha inspirado a muchos científicos a intentar imitar estos funcionamientos (Basogain, 2009).

Los intentos de crear un ordenador que sea capaz de emular estas capacidades ha dado como resultado la aparición de las llamadas Redes Neuronales Artificiales o Computación Neuronal.

Revisión histórica

Las Redes Neuronales Artificiales son una parte de la Inteligencia Artificial (IA) que sigue la corriente conexionista. Ésta hace referencia a una forma de ver la mente computacional que defiende que la información se encuentra entre los elementos que forman el sistema.

El objetivo de las Redes Neuronales Artificiales es la construcción de máquinas inteligentes con una arquitectura computacional parecida a las redes neuronales del cerebro (Montaño, 2005).

El inicio de las redes neuronales se sitúa en 1943 con la aportación de McCulloch y Pitts, creadores de la primera neurona artificial, cuya estructura se muestra en la Figura 1. Su funcionamiento era muy sencillo y consistía en que si la suma de las entradas excitatorias supera el umbral de activación de la unidad sin la presencia de una entrada inhibitoria, la

neurona se activa y emite una respuesta (equivalente al valor de 1). Pero si hay la presencia de una entrada inhibitoria, no se activará y no emitirá respuesta (representado por el valor 0).

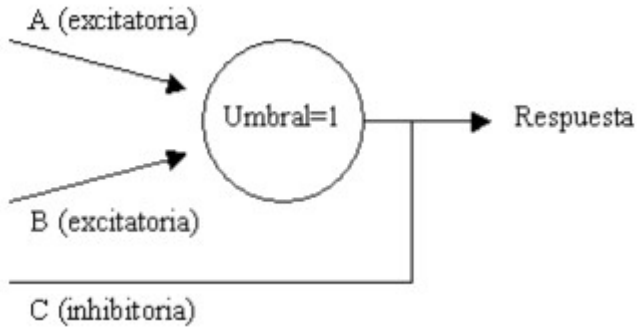


Figura 1: *Primera neurona de McCulloch y Pitts*

La segunda aportación más importante fue la de A. Hebb y su regla de aprendizaje hebbiano en 1949, la cual explica el mecanismo de regulación entre las conexiones. Este aprendizaje propone que los pesos se incrementan si las neuronas de origen y de destino están activadas simultáneamente (Díaz, Gómez, Graciani y López, 1994) reforzando esa conexión y aumentando la probabilidad de activación conjunta de ambas neuronas en el futuro. Esto quiere decir que cuanto más conjuntamente se activen las neuronas las conexiones se incrementarán.

Rosenblatt en 1958, creó la primera red neuronal denominada Perceptrón simple para el reconocimiento de patrones visuales. Éste modelo se construyó a partir de la unión de las neuronas formales de McCulloch y Pitts y el modelo Pandemonium de Selfridge (Aróztegui y Prados, 2008).

Es una red compuesta por tres capas y solo una de ellas tiene las conexiones modificables, se puede ver la estructura en la Figura 2. Era el primer modelo capaz de generalizar (reconocer

patrones similares y aprenderlos). En la capa de entrada los valores de las conexiones son fijos, son las conexiones de la capa de salida las que tiene los valores modificables, ya que les influye los valores de la capa anterior (la capa oculta).

Esta red funciona con el teorema de convergencia de la regla de aprendizaje del Perceptrón desarrollada por Rosenblatt. Ésta consiste en una corrección de errores e ir ajustando los valores de las conexiones para minimizar las diferencias entre el valor de salida obtenido y el valor de salida que se desea.

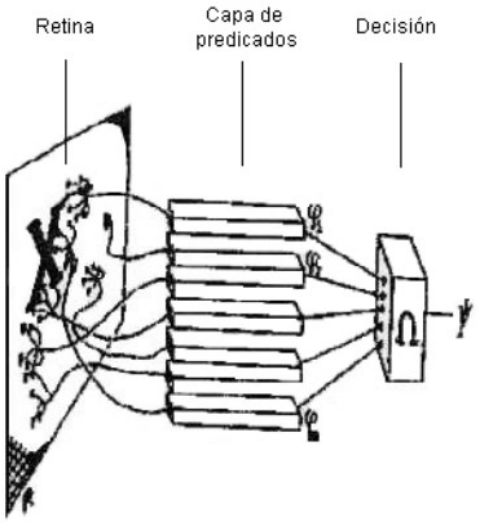


Figura 2: *Perceptrón Simple Rosenblatt*

También empezó a plantear una primera idea sobre un Perceptrón con más capas, lo que después se conocería como el Perceptrón Multicapa (MLP), pero debido a la tecnología de la época no pudo poner esa idea en práctica.

Con estas grandes aportaciones marcaron las bases para la construcción de las redes neuronales, pero en 1969 aparecen las primeras críticas a éstas y al paradigma conexionista. Uno de los problemas que presentaban las redes neuronales era a nivel tecnológico, porque

para poder construir una red hacía falta un potenciador para que funcionasen, lo cual requería demasiado espacio físico.

La crítica más importante y que más efecto tuvo sobre las redes neuronales fue la llevada a cabo por Minsky y Papert hacia el Perceptrón Simple en su libro “Perceptrons”. La principal crítica que hacían en su libro era que las redes neuronales únicamente eran capaces de resolver problemas linealmente separables, es decir, las funciones conocidas como OR. Estas redes no eran capaces de solucionar la función XOR (OR-Exclusiva), además de no poder realizar con efectividad otras funciones matemáticas, como por ejemplo, la paridad y la función topológica de la conectividad.

También señalaban la incapacidad del Perceptrón a la hora de diferenciar entre entradas de distintas partes de la pantalla, ni entre figuras con distintas posiciones de rotación, además de defender el poco uso práctico que tenían estas redes.

Aun así muchos siguieron trabajando en el avance de las redes y haciendo grandes aportaciones, como Von Der Malsburg (1973) y Fukushima (1975), Kohonen (1977), Grossberg (1980), etc. Pero la siguiente aportación que más nos interesa fue la que hicieron Rumelhart, Hinton y Williams del grupo PDP (Parallel Distributed Processing) en 1986. Esta aportación fue el algoritmo de retropropagación (backpropagation).

El objetivo del uso del algoritmo de backpropagation, es el de conseguir propagar los errores cometidos por las neuronas de salida hacia atrás. Ésto es así siguiendo la lógica de la idea de que los errores que comente una unidad intermedia depende del error cometido por las unidades de salida a las que está conectada. Una vez se conoce el error, se pueden modificar

los pesos de las conexiones de las unidades, en la dirección opuesta al gradiente, ya que así el decremento del error es más rápido.

Con todas estas aportaciones se estaba reabriendo otra vez el debate sobre el conexionismo y con ello volvieron a aparecer críticas a este movimiento. Las primeras críticas fueron encabezadas también por Minsky y Papert con una nueva edición de su libro "Perceptrons" en 1988.

Pero no fueron los únicos autores que realizaron críticas a estos nuevos trabajos. Otros científicos, líderes en sus áreas de investigación, tales como Poggio (visión), Hillis (ordenadores paralelos) y Fodor y Pylyshyn (ciencia cognitiva), también realizaron críticas radicales al conexionismo (Olazarán, 1991).

Como el tema central del trabajo es el Perceptrón Multicapa, en la revisión histórica solamente se han nombrado aquellas aportaciones más influyentes para el mismo. Pero para poder ver la historia completa de las redes neuronales resumida se puede ver en el cuadro resumen del Anexo 1.

Concepto

Las redes neuronales son sistemas de procesamiento de la información cuya estructura y funcionamiento están basados en las redes neuronales biológicas. Estos sistemas están compuestos por elementos simples denominados nodos o neuronas, los cuales se organizan por capas.

Cada neurona está conectada con otra a partir de unos enlaces denominados conexiones, cada una de las cuales tiene un peso asociado. Estos pesos son valores numéricos modificables y representan la información que será usada por la red para resolver el problema que se presente.

La actividad de esta unidad de procesamiento es sencilla. Ésta consiste en sumar los valores de entrada (inputs) que recibe de las unidades a las que se encuentra conectada. A continuación compara esta cantidad con el valor del umbral y, si lo iguala o supera, envía una salida (output) a las demás unidades con las que está conectada. En cambio, si es menor que el valor del umbral, no se produce una salida.

La arquitectura del procesamiento de la información del sistema conexionista es en paralelo, es decir, se pueden activar diversas neuronas al mismo tiempo. Otra característica del sistema conexionista es que la información no se encuentra localizada en un solo punto o compartimentos, sino que está distribuida a lo largo de las conexiones de la red.

Este tipo de distribución del conocimiento de la red tiene sus ventajas, ya que hace que la red sea tolerante a los fallos porque cada unidad lleva a cabo una computación simple y el funcionamiento de la red resulta de la intervención de todas las neuronas. Aunque una pequeña parte del sistema haya sido dañada, esto no afecta a su funcionamiento total.

Además son sistemas adaptativos que aprenden de la experiencia, es decir, que requieren de un entrenamiento para poder resolver unos problemas. Con esto nos referimos a que primero se le han de haber presentado a la red un número de situaciones parecidas a la situación problema (muestra representativa de casos). Y el aprendizaje se da cuando la red haya hecho

la representación interna del problema a partir de la modificación o ajuste de los pesos de las conexiones (autoorganización), a través de un algoritmo o regla determinados.

Con este tipo de aprendizaje la red será capaz de responder adecuadamente cuando se le presenten situaciones a las cuales no había sido expuesta con anteriormente (Montaño y Palmer, 1999), es decir, las RNA son capaces de generalizar las situaciones a casos nuevos.

Por último, la representación de la información que maneja la red, es decir, el valor de los pesos y de las entradas y salidas de la red, es numérica. Estos valores pueden ser un valor real continuo, como la edad, o un valor binario, el sexo de una persona (hombre, mujer) o respuestas de sí o no.

Fundamentos biológicos

Las redes neuronales artificiales están basadas en el funcionamiento del sistema nervioso humano, cuyo componente principal es la neurona. A continuación se presenta una breve descripción de estos.

Una neurona está compuesta por un cuerpo o soma, del cual salen unas pequeñas ramificaciones, denominadas dendritas, y una ramificación principal, el axón, el cual también se acaba ramificando en varias ramas de longitud variable, lo que constituye el terminal axónico (ver Figura 3).

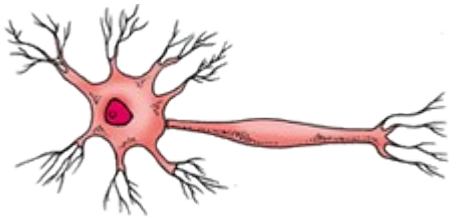


Figura 3: *Estructura de la neurona*

Las neuronas se encuentran intercomunicadas entre sí, con lo que tienen una gran capacidad comunicativa. El proceso de comunicación sigue de la siguiente forma: primero las dendritas reciben la señales de entrada y en el soma se integran y se emiten señales de salida que pasan al axón. A continuación el axón transporta las señales emitidas hasta los terminales axónicos, los cuales son los encargados de distribuir la información a las neuronas más próximas. El momento en el que la señal pasa de una neurona a la siguiente neurona se denomina sinapsis.

La señal eléctrica comienza con un cambio de permeabilidad de la membrana de la neurona, la cual deja entrar los iones del medio al interior y se da una despolarización de la neurona, lo cual constituye el potencial de acción e inicia la señal eléctrica.

Luego esta señal viaja a lo largo del axón hasta llegar al terminal sináptico, en donde la señal eléctrica se transforma en química. Cuando se llega al terminal esta señal estimula las vesículas llenas de una sustancia denominada neurotransmisores, lo cuales son liberados y viajan hasta llegar a la siguiente neurona. Una vez la alcanzan y se unen a sus receptores se crea un cambio en la siguiente neurona y se vuelve a dar la señal eléctrica.

Elementos de una red neuronal

Una red neuronal es un conjunto de unidades elementales con unas determinadas conexiones. Se pueden encontrar tres tipos de unidades o neuronas:

1. **Unidades de entrada**: son las que reciben las señales del entorno.
2. **Unidades de salida**: envían las señales fuera del sistema, es decir, dan una respuesta.
3. **Unidades ocultas**: las entradas y salidas de estas se encuentran dentro de la red, es decir, siempre conectadas con otra unidad.

Cada conjunto de neuronas cuyas señales proceden de la misma fuente y sus conexiones se dirigen al mismo destino forman una capa. Las conexiones entre las capas tienen un peso o valor asignado, el cual es importante para el aprendizaje de la red.

En función de las capas podemos encontrar dos tipos de redes. Las redes monocapa, en donde solo se presenta una única capa de neuronas, las conexiones presentes en estas redes son laterales y autorrecurentes (conexión de una unidad consigo misma). Y las redes multicapa, en las cuales existen dos o más niveles de neuronas. En cuanto a las conexiones en éstas se pueden apreciar de dos tipos: conexiones feedforward, en donde todas las señales van en una misma dirección, a la siguiente neurona. Y, por último, las conexiones feedforward y feedback, en éstas las neuronas presentan conexiones que pueden ir al siguiente nivel de neuronas y a su vez, conexiones que pueden ir al nivel anterior.

Pero también se pueden clasificar en función del tipo de asociación que hacen entre la información de entrada y la información de salida. Existen dos formas de asociación de la información lo que da lugar a dos tipos de redes:

1. **Redes Heteroasociativas**: la red aprende parejas de datos, con lo que cuando se le presenta una determinada entrada, la red tendrá que responder con su par asociado. Este tipo de redes se pueden usar para dos tareas en concreto: una de clasificación de la información o para computar una función general de su entrada.

2. **Redes Autoasociativas**: la red aprende una serie de datos y cuando se le presente una entrada será la propia red la que realice la correlación, es decir, responderá con el dato almacenado que más se parezca al dato de entrada. El objetivo del uso de este tipo de redes es el de reconstruir información que se presenta incompleta.

Mecanismos de aprendizaje

Las Redes Neuronales Artificiales aprenden a partir de la modificación de los pesos en función de una información de entrada. La forma en la que la red cambia estos valores asignados de las conexiones sigue una serie de reglas, las cuales se conocen como reglas de aprendizaje. Estas reglas se pueden clasificar de dos formas diferentes:

En función de si hay presente o no un agente externo, encontramos dos tipos de aprendizajes:

1. **Aprendizaje supervisado**: está presente un agente externo que controla el aprendizaje de la red, es el que cambia los valores para obtener la respuesta deseada.

2. **Aprendizaje no supervisado**: el agente externo no está presente, con lo que no hay control externo en el aprendizaje de la red, por esto se dice que estas redes tienen la capacidad de autoorganizarse.

Tipos de red neuronal

A continuación se presenta un cuadro resumen de todas las redes neuronales que hay organizadas según el tipo de aprendizaje y las conexiones que tienen.

1. Aprendizaje supervisado

   a) Con conexiones feedforward

      - Lineales

        Perceptrón Simple (Rosenblatt, 1958)

        ADALINE (Widrow y Hoff, 1960)

      - Perceptrón Multicapa

        Backpropagation (Hinton, Rumelhart y Williams, 1986)

        Correlación en cascada (Fahlman y Lebiere, 1989)

        Quickpropagation (Fahlman, 1988)

        Delta-bar-delta (Jacobs, 1988)

        Resilient Propagation (Braun y Riedmiller, 1993)

        Gradiente conjugado (Battiti, 1992)

      - Radial Basis Function (RBF) (Broomhead y Lowe, 1988)

        Orthogonal Least Squares (Chen, Cowan y Grant, 1991)

      - Cerebellar Articulation Controller (Albus, 1975)

      - Clasificación

        Learning Vector Quantization

        Red Neuronal Probabilística (Specht, 1990)

      - Regresión

        General Regression Neural Network (Specht, 1991)

   b) Conexiones Feedback

      - Bidirectional Associative Memory (Kosko, 1992)

      - Máquina de Boltzman (Ackley, Hinton y Sejnowski, 1985)

      - Series temporales recurrentes

Backpropagation through time (Werbos, 1990)

Elman (Elman, 1990)

Recurrent backpropagation (Pineda, 1989)

Time Delay NN (Hinton, Lang y Waibel, 1990)

c) Competitivo

- ARTMAP (Carpenter, Grossberg y Reynolds, 1991)
- Fuzzy ARTMAP (Carpenter, Grossberg, Markuzon, Reynolds y Rosen, 1992)
- Gaussian ARTMAP (Williamson, 1996)
- Neocognitrón (Fukushima, 1980)

2. No supervisado

a) Competitivo

- Vector Quantization

  Grossberg (Grossberg, 1976)

  Conscience (Desieno, 1988)

- Mapa Auto-Organizado (Kohonen, 1982)
- Teoría de la Resonancia Adaptativa (ART)

  ART 1 (Carpenter y Grossberg, 1987a)

  ART 2 (Carpenter y Grossberg, 1987b)

  ART 2-A (Carpenter, Grossberg y Rosen 1991a)

  ART 3 (Carpenter y Grossberg, 1990)

  Fuzzy ART (Carpenter, Grossberg y Rosen, 1991b)

  Differential Competitive Learning (Kosko, 1992)

b) Reducción de dimensionalidad

- Regla de Oja (Oja, 1992)

- Sanger (Sanger, 1989)
- Differential hebbian (Kosko, 1992)

c) Autoasociación

- Autoasociador lineal (Anderson, Silverstein, Ritz y Jones, 1977)
- Brain-State-in-a-Box (BSB) (Anderson, Silverstein, Ritz y Jones, 1977)
- Red de Hopfield (1982)

Perceptrón Multicapa

Pero de todas estas redes únicamente explicaremos el Perceptrón Multicapa junto con su algoritmo de *backpropagation error*, ya que es la red que se va ha usar en este trabajo y ya que es la red la más usada actualmente.

El porque esta red es la más usada actualmente se debe principalmente a que es capaz de actuar como aproximador universal de funciones. Además junto con el algoritmo de *backpropagation* es capaz de aprender cualquier tipo de función continua entre un grupo de variables de entrada y de salida (Jiménez, Montaño y Palmer, 2001).

Dependiendo de la complejidad de la red, el Perceptrón puede resolver desde funciones continuas, si únicamente presenta una sola capa oculta, hasta funciones no continuas, si presenta más de una capa oculta puede resolver las funciones discontinuas.

El uso del algoritmo backpropagation permite que la red aprenda la asociación existente entre un conjunto de patrones de entrada y sus salidas. Esto se consigue a partir de una representación interna del conocimiento en las capas ocultas a fin de aprender la relación y

aplicar esta misma a las nuevas entradas. En pocas palabras, permite generalizar la información.

Esta red está compuesta por una capa de entrada, una de salida y una o más capas ocultas. En la Figura 4 se muestra la estructura de un Perceptrón Multicapa simple, es decir, con una sola capa oculta. Las conexiones son siempre feedforward. El tipo de aprendizaje es supervisado, es decir, que es el usuario quien determina la salida deseada.

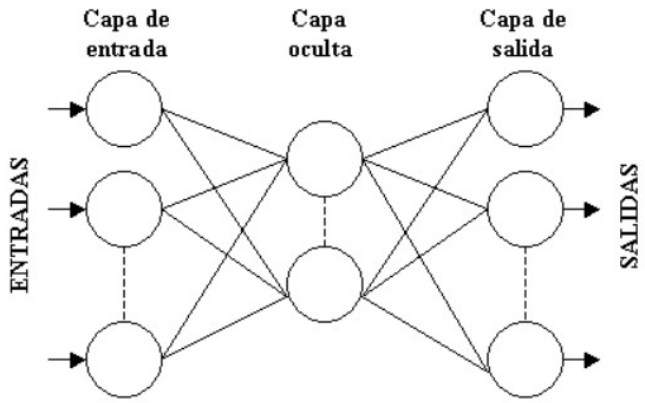


Figura 4: *Perceptrón Multicapa simple*

El algoritmo usado, es el nombrado anteriormente, backpropagation error. Para que la red pueda aprender y adquiera esa capacidad de generalizar se diferencian dos etapas: de entrenamiento y de funcionamiento.

La primera etapa de entrenamiento consiste en la presentación de un grupo denominado grupo de entrenamiento en el cual, cuyo principal objetivo es conseguir la mínima discrepancia entre el valor de salida obtenido y el valor de salida deseado.

En la segunda etapa, la etapa de funcionamiento, se presenta un nuevo grupo de datos que no han sido presentados anteriormente, grupo de validación, la señal se transmite hacia las neuronas ocultas en donde se transforman las señales a partir de la aplicación de una función

de activación y proporcionan un valor de salida que se transmite a la siguiente capa, en donde, a partir de la misma operación se obtiene una salida de la red. El objetivo de este grupo es el de controlar el proceso de aprendizaje.

Para ver un funcionamiento más detallado del Perceptrón Multicapa ver el Anexo 2.

Aplicaciones de las Redes Neuronales Artificiales

Como se ha ido viendo a lo largo del trabajo las RNA tienen diferentes aplicaciones en distintos campos, como puede ser en biología para el estudio del cerebro u obtención de modelos de retina, en el medio ambiente, para la previsión del tiempo, en finanzas, previsión de la evolución de los precios, identificación de firmas, aplicaciones militares, en la clasificación de las señales de radar, creación de armas inteligentes, detección de bombas, etc.

Pero no hablaremos de todos ellos, sino que únicamente nos centraremos en las distintas áreas del campo de la Psicología ya que es el área de interés de este trabajo. A partir de la revisión de diferentes artículos se ha encontrado la aplicación de las redes neuronales en al ámbito de la Psicología de la Educación para la predicción del rendimiento académico o en la detección de problemas de aprendizaje, en esta se ha aplicado modelos de redes neuronales para poder entender mejor los procesos didácticos. Un ejemplo de esta aplicación es la llevada a cabo en un trabajo de investigación de Cataldi, Lage y Salgueiro (2007), los cuales crearon una red neuronal para predecir, a partir del desempeño de los estudiantes, si aprobarían o no una determinada asignatura. También se han aplicado para ayudar en el aprendizaje de niños con dificultades en este, como por ejemplo, dislexia, hiperactividad (TDAH), etc.

En el área de Recursos Humanos, se puede encontrar aplicaciones de las RNA en la selección de personal. El uso en esta área tiene como finalidad agilizar el proceso de selección del personal a través de la extracción de diversas características sociológicas de los candidatos a un puesto, como por ejemplo sería, la personalidad. Un ejemplo de esto fue un estudio llevado a cabo por Acevedo, Caicedo y Loaiza (2007) para la selección de personal para la Armada Nacional. En este crearon una red del tipo Perceptrón Multicapa y otro de base radial. En este usaron una base de datos creadas a partir de los test de personalidad pasados a los candidatos. La arquitectura de la red Perceptrón constaba de 17 neuronas de entrada, 18 neuronas ocultas con función sigmoidal y 1 neurona de salida con función lineal.

En el área de la Psicología Social se han aplicado las redes neuronales en el análisis de la conducta y crear modelos de redes basados en los dos elementos más importantes que dirigen la conducta humana, el reforzamiento y castigo.

En otros trabajos sobre esta misma área, por ejemplo el de Bringas, Herrero, Ovejero y Rodríguez (2008), usaron las redes neuronales para predecir el valor predictivo de las variables de personalidad, motivación y uso de los aparatos electrónicos sobre las conductas antisociales en adolescentes.

En cuanto al área de la Psicología clínica, se han desarrollado redes neuronales para la detección de diversos trastornos como la ansiedad, depresión postparto, etc. Un ejemplo es una aplicación desarrollada para móviles que permite detectar la depresión postparto a partir de cuestionarios pasados a las madres tras éste (Jiménez, 2013).

En Neurología también se han aplicado estas redes para hacer una análisis estadístico de las imágenes tomográficas cerebrales para poder hacer una mejor detección de los casos de Alzheimer en las personas mayores.

Las Redes Neuronales Artificiales han hecho grandes aportaciones en el campo de los procesos psicológicos básicos como es la adquisición del lenguaje. Una aportación interesante es la de Betancur, Peña y Vélez (2013) es el hacer uso de las redes neuronales para la traducción automática del lenguaje dactilológico de sordos y sordomudos mediante sistemas adaptativos, en el cual, por ahora, ya han conseguido traducir con éxito las vocales del vocabulario español,

Uso de las Redes Neuronales Artificiales en las Conductas Adictivas

Las conductas adictivas son unas conductas muy complejas ya que dependen de una gran cantidad de factores conocidos y desconocidos, con lo que la relación que se establece entre estos y la conducta no se puede reducir a una relación directa de causa y efecto.

Las Redes Neuronales Artificiales aplicadas en este campo han resultado ser de gran utilidad. Hay diversos estudios que se han llevado a cabo en este campo que comprueban la eficacia del uso de las redes neuronales en la predicción de conductas adictivas y de elección de tratamiento. Un ejemplo de este hecho es el estudio llevado a cabo por Buscema, Caricati, Grassi, Intraligi y Necini (2001) en donde se comparaban las decisiones de las Redes Neuronales Artificiales con las decisiones de los profesionales sobre qué tratamiento aplicar en casos de drogadicciones mostraron que los tratamientos que eran elegidos por las redes tenían mejores predicciones (62,7%).

En otros estudios llevados a cabo sobre las adicciones es el consumo de alcohol en adolescentes (Cajal, Gervilla, Jiménez, Montaño, Palmer y Sesé, 2008) en donde se comparaban tres tipos de modelos para el análisis de datos se encontró que las redes neuronales eran el modelo más preciso, tenían una tasa de acierto del 64,1%.

Sobre el consumo de tabaco un grupo de investigadores aplicaron una red neuronal con el algoritmo de *backpropagation* para ver la capacidad de clasificación sobre una muestra de fumadores y antiguos fumadores. En este estudio los resultados no fueron tan prometedores como en los nombrados anteriormente, ya que la red no fue capaz de hacer una clasificación con éxito (McDaniel y Poynton, 2006).

Objetivo

El principal objetivo de este trabajo es demostrar el uso del Perceptrón Multicapa en la predicción en las conductas adictivas de sujetos en tratamiento. En concreto, el objetivo es predecir el número de sujetos que han recaído en 5 meses y medio.

Metodología

Matriz de datos

La matriz de datos que se va ha utilizar para llevar a cabo este trabajo proceden de una serie de estudios realizados por el equipo de McCusker (McCusker et al., 1995; McCusker, Bigelow, Frost et al., 1997; McCusker, Bigelow, Vickers-Lahti, Spotts, Garfield y Frost, 1997) en la Universidad de Massachusetts.

El objetivo de estos estudios era el de comparar diversos programas de intervención, que se diferenciaban en función de la duración de la intervención (corta o larga) y de la orientación terapéutica (clínica A o B), para conseguir la reducción del abuso de drogas en 628 toxicómanos.

En la siguiente tabla se presentan las diferentes variables tenidas en cuentas en los estudios. La variable respuesta que se utiliza es si después de 5 meses y medio es si el sujeto recae o no en el consumo de drogas.

Tabla 1

Descripción de las variables explicativas

Variable	**Descripción**	**Valores**
Edad	Edad al entrar en el estudio	años
Beck	Puntuación en el Inventario de Depresión de Beck	0 - 54
Hercoc	Uso de heroína/cocaína 3 meses antes de la admisión	1 = heroína y cocaína; 2 = solo heroína; 3 = solo cocaina; 4 = ni heroína, ni cocaína
Hdroga	Historia del uso de droga	1 = nunca; 2 = anterior; 3 = reciente
Ntrat	Número de tratamientos previos contra la droga	0 - 40
Raza	Raza del sujeto	0 = blanca; 1 = no blanca
Trat	Asignación aleatoria de tratamiento	0 = corto; 1 = largo
Lugar	Lugar del tratamiento	0 = clínica A; 1 = clínica B
Recaída	Si el sujeto recae o no recae en el consumo de drogas a los 5 meses y medio (166 días) de seguimiento	0 = recae; 1= no recae

El momento temporal escogido es muy importante ya que se eligió el punto en donde la mitad de los participantes recaían en el consumo 315 y la otra mitad no, 313. Esto es una ventaja a

la hora de comprobar el funcionamiento de la red y saber si la clasificación que se ha hecho es correcta.

Tabla 2.

Supervivencia

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Recae	315	50,2	50,2	50,2
	No recae	313	49,8	49,8	100,0
	Total	628	100,0	100,0	

En la siguiente tabla se muesta los porcentajes de sujetos presentes en cada uno de los grupos que se formaron para poner a prueba el funcionamiento del modelo. Los participantes se asginaron de forma aleatoria a cada uno de los grupos. El grupo de mayor tamaño es el de entrenamiento con 314 particpantes, y los grupos validación y test presentan 157 participantes, tal y como se muestra en la siguiente tabla.

Tabla 3

Grupo

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Entrenamiento	314	50,0	50,0	50,0
	Validación	157	25,0	25,0	75,0
	Test	157	25,0	25,0	100,0
	Total	628	100,0	100,0	

Teniendo el mimo momento temporal, en cada uno de los grupos la mitad de los participantes recaía en el consumo de drogas y la otra mitad no. Esto se muestra en la siguiente tabla, donde se muestra la cantidad de participantes que reacaen y los que no.

Tabla 4.

Contingencia Grupo * Supervivencia

Recuento		Supervivencia		Total
		Recae	No recae	
Grupo	Entrenamiento	158	156	314
	Validación	79	78	157
	Test	78	79	157
Total		315	313	628

Instrumentos

El instrumento principal de este trabajo es la aplicación *Sensitive Neuronal Network* (SNN) creada por Fernández, Montaño y Palmer en 2001. Ésta se usó para la creación del modelo de red neuronal más eficaz para hacer la clasificación de la muestra.

Programa SPSS para la realización de la curva de ROC y así poder analizar la sensibilidad del modelo escogido. Además también se usó el programa para la creación de las tablas.

Procedimiento

Los pasos seguidos para el diseño del modelo de red neuronal han sido los siguientes: selección de las variables relevantes, procesamiento de los datos, creación de los grupos y selección y construcción del modelo de red neuronal.

Selección de las variables: las variables relevantes usadas para confeccionar la muestra han sido: edad, las puntuaciones en el Inventario de Depresión de Beck, el consumo de heroína 3 meses antes de la admisión en el programa, historia del uso de la droga, número de tratamientos previos, raza del participante, tratamiento asignado aleatoriamente, lugar de tratamiento y recaída del sujeto.

Son nueve variables en total, con lo que se usarán nueve neuronas de entrada y una neurona de salida, ya que la respuesta que se busca es si el sujeto recae en el consumo de drogas o no después de haber pasado cinco meses y medio.

La función usada en las neuronas de entrada es una función lineal. En cambio, en la neurona de salida se aplicará una función sigmoidal ya que se realiza una tarea de predicción y para estas tareas se recomienda el uso de esta función ya que los valores no son continuos, sino que dentro de los valores dados (0 y 1), la respuesta se puede desplazar de forma no contínua en ellos.

Procesamiento de los datos: los valores de la muestra obtenida del estudio que llevó a cabo el equipo de McCusker se transformaron a valores comprendidos entre -1 y 1. Así pues, los valores de las nueve variables que se han tenido en cuenta a la hora de conformar la muestra se transformaron para establecerla en una nueva distribución comprendida entre los valores nombrados anteriormente.

Al haber variables continuas, como por ejemplo, la edad, se usó una fórmula, que se muestra en la figura 5 y 6, para transformar los valores.

$$A = r(V - V_{min}) + A_{min}$$

Figura 5: Expresión usada para la transformación de los datos

$$r = \frac{A_{max} - A_{min}}{V_{max} - V_{min}}$$

Figura 6: Expresión usada para encontrar el valor de r

Creación de los grupos de entrenamiento, validación y test: la muestra de 628 sujetos se ha dividido en tres grupos: entrenamiento, el cual se usará en la fase de aprendizaje, validación, usado en la fase de funcionamiento, y test, aplicado en la etapa test para demostrar el funcionamiento de la red.

Para ello, el grupo de entrenamiento consta de 314 sujetos para que la red sea capaz de ajustar los pesos de forma iterativa. Luego el grupo de validación y test están compuestos por 157 sujetos cada uno. El grupo de validación ha servido para averiguar cual el número de pesos óptimo y escoger el modelo de red que ha tenido un mejor desempeño con este grupo.

*Selección y construcción del modelo de red neurona*l: se usará el modelo de red el Perceptrón Multicapa, ya que es la red de interés del trabajo y que además se ajusta al objetivo del mismo, el cual es la clasificación de un grupo de sujetos con conductas adictivas en si recaen o no en el consumo de drogas.

Para encontrar el modelo más eficaz he ido creando diferentes modelos modificando los siguientes valores: número de neuronas ocultas, tasa de aprendizaje y factor momento. El punto de corte que se ha seleccionado y que se ha mantenido en todos los modelos es 0,5.

El primer valor en modificarse ha sido el número de neuronas ocultas para encontrar el modelo con el mejor desempeño, el número de neuronas que se han aplicado han sido 1, 2, 3 y 4. Después se ha modificado el valor de la tasa de aprendizaje y se ha mirado los valores 0,05, 0,1 y 0,2. Y, por último, se modificó el factor momento, dejando los otros valores estables, por los valores 0,8, 0,9 y 1.

En total, he creado 14 modelos de red neuronal, para escoger entre todos ellos primero se ha eliminado aquellos que no tuvieran un funcionamiento estable, como sucedía en los modelos 10 y 13. A continuación , he mirado el que tuviera un desempeño más equilibrado y mayor en las fases de entrenamiento y validación, para que su desempeño en la fase de test sea más estable.

Siguiendo estos criterios el modelo que se seleccionó para pasar a la fase test fue el modelo 1, ya que aunque el modelo 8 tenía un mayor desempeño en la fases de entrenamiento había mucha diferencia de porcentaje de acierto con la fase de validación. Una vez seleccionado el modelo, se pasó a la fase final, la fase test, para obtener el rendimiento real de la red.

Si se quieren ver los valores que se han obtenido en los diferentes modelos que se han creado, se puede consultar el Anexo 3, en donde se encuentran las tablas anexo 1 y 2 con los valores.

Evaluación del rendimiento del modelo: el análisis se hizo a través de la curva ROC obtenida a través del prgrama SPSS. Se usó este método de análisis ya que nos encontramos ante un

problema de clasificación de dos variables y este es la medida precisa y válida para evaluar la precisión diagnóstica del modelo seleccionado.

Resultados

A continuación presentaremos los resultados obtenidos de las clasificaciones en cada uno de los grupos creados junto con el análisis de sensibilidad realizado con la curva ROC.

Grupo de entrenamiento

El grupo de entrenamiento, formado por 314 sujetos, se obtuvo que 181 participantes recaían y 133 no. Pero de esta clasificación, la red consiguió clasificar correctamente el 84.39% de los patrones. De los 181 participantes 145 son los correcto y 36 se encuentran mal discriminados. Al igual que en en los patrones de no recaída, en donde 120 están correctamente clasificados y 13 no lo están. Estos datos se encuentran resumidos en la siguiente tabla.

Tabla 5

Tabla de contingencia Valor observado * Valor estimado

Recuento

		Valor estimado		Total
		Recae	No recae	
Valor observado	Recae	145	13	158
	No recae	36	120	156
Total		181	133	314

A estos resultados se analizaron a través de la curva ROC y se obtuvo que el área bajo la curva es de 0.855. El cual es un valor proximo a uno, así pues el modelo tiene una buena

capacidad de clasificación. Además la hipótesis obtenida es alternativa, con lo que los resultados son significativos para este grupo de clasificación.

A continuación se muestra la tabla y el gráfico de los resultados de la curva ROC, que respaldan lo nombrado anteriormente.

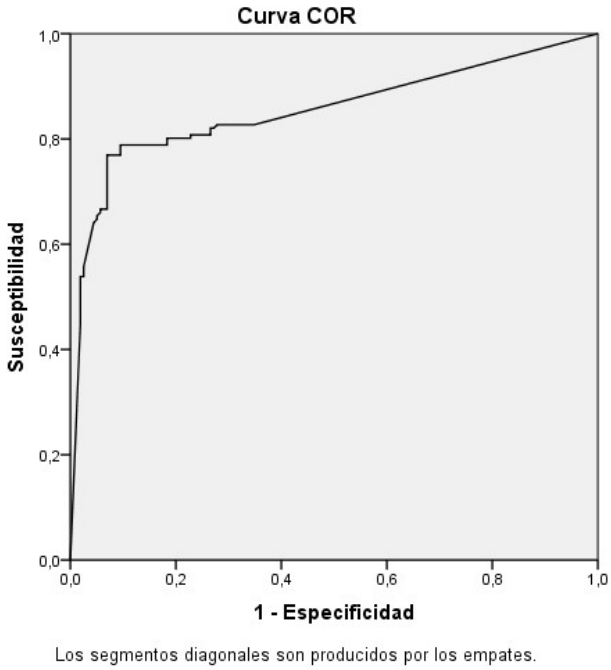


Gráfico 1: Curva ROC entrenamiento

Tabla 6

Área bajo la curva				
Variables resultado de contraste: Valor estimado				
Área	Error típ.[a]	Sig. asintótica[b]	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
,855	,023	,000	,810	,900
La variable (o variables) de resultado de contraste: Valor estimado tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Los estadísticos pueden estar sesgados .				
a. Bajo el supuesto no paramétrico				
b. Hipótesis nula: área verdadera = 0,5				

Grupo de validación

En cuanto al grupo de validación, foramdo por 157 participantes, de los cuales 85 han sido clasificados como recaídas y 72 como no recaída. Del grupo de recaída 67 son verdaderos positivos y 18 falsos positivos. En el grupo de no recaída 60 se encuentran correctamente clasificados y 12 no. Por lo tanto, el 80.89% de los patrones se encuentran bien clasificados. En la tabla 7 se muestra un resumen de los datos obtenidos.

Tabla 7

Tabla de contingencia Valor observado * Valor estimado				
Recuento				
		Valor estimado		Total
		Recae	No recae	
Valor observado	Recae	67	12	79
	No recae	18	60	78
Total		85	72	157

La curva ROC muestra que el área bajo la curva es de 0.793. El cual es un valor proximo a uno, así pues el modelo tiene una buena capacidad de clasificación. Además la hipótesis obtenida es alternativa, con lo que los resultados son significativos para este grupo de clasificación.

A continuación se muestra la tabla y el gráfico de los resultados de la curva ROC, que respaldan lo nombrado anteriormente.

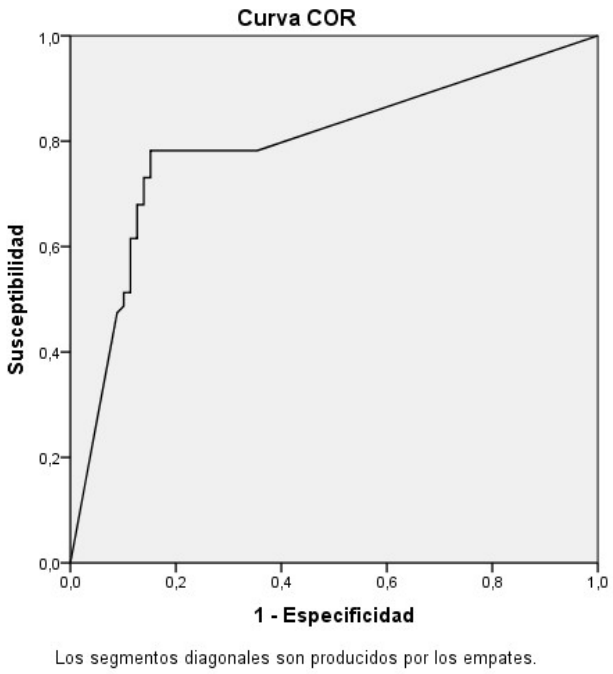


Gráfico 2: Curva ROC validación

Tabla 8

Área bajo la curva				
Variables resultado de contraste: Valor estimado				
Área	Error típ.[a]	Sig. asintótica[b]	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
,793	,038	,000	,719	,867
La variable (o variables) de resultado de contraste: Valor estimado tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Los estadísticos pueden estar sesgados .				
a. Bajo el supuesto no paramétrico				
b. Hipótesis nula: área verdadera = 0,5				

Grupo test

En cuanto al grupo de test (157 participantes), el cual es el que nos interesa ya que muestra el verdadero rendimiento de la red, los resultados obtenidos son que 88 (56,05%) de los participantes del grupo test recaen en las drogas y 69 (43,95%) no recaen.

De estos resultados, la red ha conseguido predecir correctamente 117 (74,5%) patrones. De los 117 patrones predecidos 63 de ellos son de recaída y 54 de no recaída. De los 40 patrones mal predecidos, 25 son de recaída y 15 de no recaída.

Tabla 9

Tabla de contingencia Valor observado * Valor estimado				
Recuento				
		Valor esimado		Total
		Recae	No recae	
Valor observado	Recae	63	15	78
	No recae	25	54	79
Total		88	69	157

La curva ROC muestra que el área bajo la curva es de 0.78. El cual es un valor muy parecido al obtenido en el grupo de validación y está proximo a uno, así pues el modelo tiene una buena capacidad de clasificación. Además la hipótesis obtenida es alternativa, con lo que los resultados son significativos para este grupo de clasificación.

A continuación se muestra la tabla y el gráfico de los resultados de la curva ROC, que respaldan lo nombrado anteriormente.

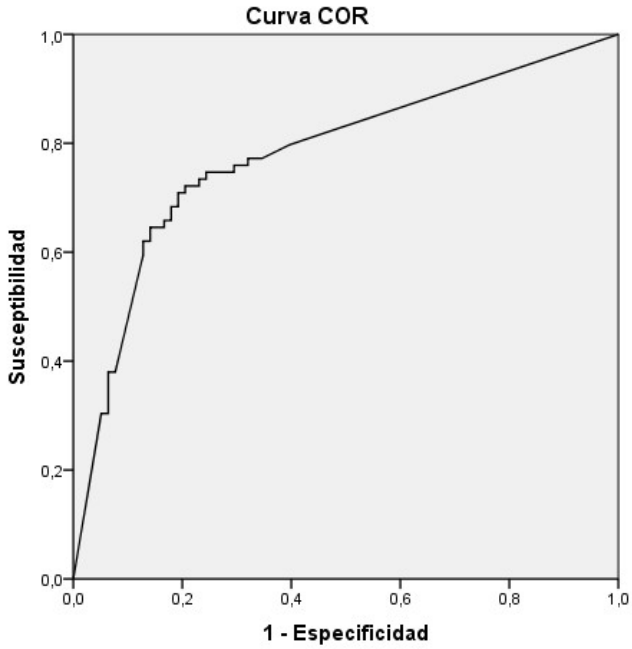


Gráfico 3: Curva ROC test

Tabla 10

Área bajo la curva				
Variables resultado de contraste: Valor estimado				
Área	Error típ.[a]	Sig. asintótica[b]	Intervalo de confianza asintótico al 95%	
			Límite inferior	Límite superior
,780	,038	,000	,705	,854
La variable (o variables) de resultado de contraste: Valor estimado tiene al menos un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Los estadísticos pueden estar sesgados .				
a. Bajo el supuesto no paramétrico				
b. Hipótesis nula: área verdadera = 0,5				

Conclusión

En este trabajo se ha querido mostrar una red neuronal, el Perceptrón Multicapa, capaz de de clasificar a los participantes según si recaían o no en el consumo de drogas, en concreto, en el consumo de heroína.

La eficacia del modelo seleccionado es del 74,5 %, el cual es superior a los diversos estudios encontrados sobre el consumo de drogas en donde la eficacia de las redes no era superior al 70%. Respaldando los resultado obtenidos el análisis de sensibilidad realizado con la curva ROC muestra que el modelo escogido tiene una gran capacidad de discriminación entre recaídos y no recaídos. Aun así se puede concluir que el Perceptrón Multicapa es un tipo de red capaz de realizar clasificación sobre las conductas adictivas de una forma eficaz.

Los modelos estadísticos clásicos, se creía que estos dan prioridad a la posibilidad de generalizar los datos, a diferencia de IA, la cual busca ofrecer soluciones algorítmocas con un coste bajo. En comparación de ambos modelos, el uso de las RNA para el análisis de datos es más fácil ya que no es necesario que se tengan en cuenta los diferentes supuestos estadísticos: homogeneidad de la matriz, normalidad, mal procesamiento de los datos, etc. Además de modificar los pesos de las conexiones de forma automática y de ser un sistema tolerante a los fallos.

Bibliografía

Acebedo, G.L.; Caicedo, E.F. y Loaiza, H. (2007). Selección de personal mediantes redes neuronales artificiales. *Revista de Matemática. Teoría y Aplicaciones*, *14*(1), doi: http://dx.doi.org/10.15517/rmta.v14i1.278

Ackley, D.H., Hinton, G.E. y Sejnowski, T.J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9*, 147-169.

Albus, J.S. (1975). New approach to manipulator control: The Cerebellar Model Articulation Controller (CMAC). *Transactions of the ASME Journal of Dynamic Systems, Measurement, and Control*, *97*(3), 220-227.

Algarabel, S.; Dasi, C.; Pitarque, A.; Ruiz, J.C. y Soler, M.J. (1993). Estudio histórico de la influencia del conexionismo en la ciencia cognitiva actual. *Revista de Historia de la Psicología,* URL: http://u.jimdo.com/www400/o/sfcb11628de7748e1/download/m52c029b00246479e/1392330050/05.+RUIZ.pdf

Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. *QÜESTIIÓ*, *25*(3), 479-498.

Anderson, C.W.; Barto, A.G. y Suttor, R.S. (1983). Neuronlike Adaptative Elements That Can Solve Difficult Learning Control Problems. I*EEE Transactions on Systems, Man and Cybernetics, 13*(5), pp. 835-846

Anderson, J.A., Jones, R.S., Ritz, S.A. y Silverstein, J.W. (1977). Distinctive features, categorical perception and probability learning: some applications of a neural model. *Psychological Review*, *84*, 413-451.

Arias, A. (2012). Avatares del paradigma conexionista. *Ciencia Cognitiva*, URL: http://medina-psicologia.ugr.es/cienciacognitiva/?p=430

Aróztegui, J. y Prados, J.M. (2008). La perspectiva conexionista: hitos y limitaciones. *Revista de historia de la psicología*, URL: http://dialnet.unirioja.es/descarga/articulo/2749336.pdf

Basogain, X. (2009). Redes Neuronales Artificiales y sus Aplicaciones. Escuela superior de Ingeniería de Bilbao, Departamento de Ingeniería de Sistemas y Automática. URL: http://www.ciberesquina.una.edu.ve:8080/2014_2/350_E.pdf

Battiti, R. (1992). First and second order methods for learning: between steepest descent and Newton's method. *Neural Computation*, *4*(2), 141-166.

Bertona, L.F. (2005). Entrenamiento de redes neuronales basado en algoritmos evolutivos. Facultad de Ingeniería, Universidad de Buenos Aires, 11-14

Betancur, D.; Peña, A. y Vélez, M. (2013). Traducción automática del lenguaje dactilológico de sordos y sordomudos mediante sistemas adaptativos. *Revista Ingienería Biomédica*, URL: http://repository.eia.edu.co/bitstream/11190/517/1/RBI00107.pdf

Braun, H. y Riedmiller, M. (1993). A direct adaptive method for faster backpropagation learning: the Rprop algorithm. *IEEE*, 586-591.

Bringas, C.; Herrero, F.J.; Ovejero, A. y Rodríguez, F.J. (2008). Medios electrónicos y comportamiento antisocial en adolescentes. *Revista Colombiana de Psicología*, *17*(1), 93-104.

Broomhead, D.S. y Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems*, *2*, 321-355.

Buscema, M.; Caricati, A.M.; Grassi, M.C.; Intraligi, M. y Necini, P. (2001). Artificial neural network assessment of substitutive pharmacological treatments in hospitalised intravenous drug users. *Artificial Intelligence in Medicine*, doi: http://dx.doi.org/10.1016/S0933-3657(01)00093-8

Cajal, B.; Gervilla, E.; Jiménez, R.; Montaño, J.J.; Palmer, A. y Sesé, A. (2009). La metodología del Data Mining. Una aplicación al consumo de alcohol en adolescentes. *Adicciones*, *21*(1), 65-80.

Calafat, A.; Montaño, J.J. y Palmer, A. (2000). Predicción del consumo de éxtasis a partir de redes neuronales artificiales. *Adicciones*, *12*(1), 29-41

Campanario, J. M. (2004). El enfoque conexionista en psicología cognitiva y algunas aplicaciones sencillas en didáctica de las ciencias. *Enseñanza de las Ciencias, 22*, 093-104

Carpenter, G.A. y Grossberg, S. (1987a). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, *37*, 54-115.

Carpenter, G.A. y Grossberg, S. (1987b). ART2: Self-organization of stable category recognition codes for analog input patterns. *Applied Optics*, *26*(23), 4919-4930.

Carpenter, G.A. y Grossberg, S. (1990). ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, *3*(4), 129-152.

Carpenter, G.A.; Grossberg, S.; Markuzon, N.; Reynolds, J.H. y Rosen, D.B. (1992). Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, *3*(5), 698-713.

Carpenter, G.A., Grossberg, S. y Rosen, D.B. (1991a). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, *4*, 493-504.

Carpenter, G.A., Grossberg, S. y Rosen, D.B. (1991b). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, *4*, 759- 771.

Cataldi, Z.; Lage, F.J. y Salgueiro, F. (2007). Predicción del rendimiento de los estudiantes y diagnóstico usando las redes neuronales. XIII Jornadas de Enseñanza Universitaria de la Informática, 187-194

Chen, S.; Cowan, C.F.N. y Grant, P.M. (1991). Orthogonal least squares learning for radial basis function networks. *IEEE Transactions on Neural Networks*, *2*(2), 302-309.

Córdoba, F. (2005). Fundamentos biológicos del aprendizaje y la memoria. Universidad de Huelva, Departamento de Biologia ambiental y Salud Pública, URL: http://www.uhu.es/francisco.cordoba/asignaturas/FBAM/TEMAS%20PDF/7-PLASTICIDAD.pdf

Desieno, D. (1988). Adding a conscience to competitive learning. Proceedings of the International Conference on Neural Networks, I, 117-124.

Díaz, M.A.; Gómez, F.J.; Graciani M.A. y López, M.T. (1994). Aprendizaje con Redes Neuronales Artificiales. *Dialnet*, URL: dialnet.unirioja.es/descarga/articulo/2281678.pdf

Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Escobar, R. (2014). Redes neuronales, procesos cognoscitivos y análisis de la conducta. Conducta. *Revista Internacional de Interconductismo y Análisis de conducta*, URL: http://conductual.com/sites/default/files/pdf-articles/Redes%20neuronales_Escobar_0.pdf

Fahlman, S.E. (1988). Faster-learning variations on back-propagation: an empirical study. Carnegie Mellon University, Computer Science Department.

Fahlman, S.E. y Lebiere, C. (1989). The cascade-correlation learning architecture.Carnegie Mellon University, Computer Science Department.

Fukushima, K. (1980). Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, URL: https://www.cs.princeton.edu/courses/archive/spring08/cos598B/Readings/Fukushima

1980.pdf

García, P. (2004). Introducción a las Redes Neuronales y su Aplicación a la Investigación Astrofísica. Instituto de Astrofísica de Canarias (IAC). URL: http://www.iac.es/sieinvens/SINFIN/Sie_Courses_PDFs/NNets/confiac.pdf

Geltatt, C.D.; Kirpatrick, S. y Vecchi, M.P. (1983). Optimization by Simulated Annealing. Science. *New Series*, *220*(4598), 671-680.

Gento, A.M. y del Olmo, R. (1994). Redes neuronales Fuzzy en la gestión de empresas. *Dialnet*, URL: dialnet.unirioja.es/descarga/articulo/116388.pdf

Gestal, M. Introducción a las Redes de Neuronas Artificiales. Dpto. Tecnologías de la Información y las Comunicaciones. Universidad de Coruña. URL: http://sabia.tic.udc.es/mgestal/cv/RNAtutorial/TutorialRNA.pdf

Gonzales, D.L; Palacio, A.M. y Piñeros, J.E. (2013). Formulación de un modelo teórico de aplicaciones de las redes neuronales en la solución de problemas de dislexia. Facultad de Ingeniería, Universidad tecnologica de Pereira.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics*, *23*, 121-134.

Gupta, P. (2012). The Cognitron and Neocognitron notes. *Hacks and Engineering*, URL: https://sites.google.com/site/hacksbyredbios/Downhome/Topic5/thecognitronandneo cognitronnotes

Henderson, H. (2007). Artificial Intelligence: Mirrors for the Mind. URL: https://books.google.es/books?id=vKmIiICDIwgC&pg=PA59&lpg=PA59&dq=SNARC+Neuman+Minsky&source=bl&ots=nPCzPnHuj0&sig=0hpNbLAswv6yOJvtPyq5o3S9HpU&hl=es&sa=X&ei=hMo3VZv3J4XT7Qbc7oHoCg&ved=0CFAQ6AEwBg#v=onepage&q=SNARC%20Neuman%20Minsky&f=false

Hilera, J.R. y Martínez, V.J. (1995). Redes Neuronales Artificiales. Fundamentos, Modelos y Aplicaciones. Madrid, RA-MA, Editorial.

Hinton, G.; Lang, K.J. y Waibel, A.H. (1990). A time-delay neural network architecture for isolated word recognition. *Neural Networks*, *3*, 23-44.

Hinton, G.E.; Rumelhart, D.E. y Williams, R.J. (1986). Learning internal representations by error propagation. Basic Mechanisms, URL: http://psych.stanford.edu/~jlm/papers/PDP/Volume%201/Chap8_PDP86.pdf

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, *79*, 2554-2558.

Hoff, M.E. y Widrow, B. (1960). Adaptative Switching Circuits. The Wescon Convention Record, 4, 98-104.

Jacobs, R.A. (1988). Increased rates of convergence through learning rate adaptation. *Neural Networks*, *1*(4), 295-308.

Jiménez, S. (2013). Desarrollo de modelos predictivos y una aplicación móvil para la predicción de la depresión postparto. Universidad de Valencia, URL: https://riunet.upv.es/bitstream/handle/10251/37483/memoria-TFM-MIB-PrediccionDPP-Santiago_Jimenez.pdf?sequence=1&isAllowed=y

Jiménez, R.; Montaño, J.J y Palmer, A. (2001). Tutorial sobre Redes Neuronales Artificiales: el perceptrón multicapa. *Psiquiatria*, URL: http://www.psiquiatria.com/psiq_general_y_otras_areas/investigacion-86/metodologia/estadistica/tutorial-sobre-redes-neuronales-artificiales-el-perceptron-multicapa/

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*, 59-69.

Kosko, B. (1992). Fuzzy Sistems as Universal Approximators. *IEER*, 1153-1162.

Lahoz-Beltra, R. (2004). Bioinformática: Simulación, vida artificial e inteligencia artificial. Madrid, Díaz de Santos, S.A., 430-434.

McDAniel, A.M. y Poynton, M.R. (2006). Classification of smoking cessation status with a backpropagation neural network. *Journal of Biomedical Informatics*, doi: http://dx.doi.org/10.1016/j.jbi.2006.02.016

Martín, J.D.; Serrano, A.J. y Soria, E. (2010). Redes Neuronales Artificiales (Trabajo de Doctorado). Universidad de Valencia, Departamento de Ingeniería. Valencia.

Matich, D.J. (2001). Redes Neuronales: Conceptos Básicos y Aplicaciones. Universidad tecnológica Nacional, Departamento de Ingeniería Química. URL: http://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monograias/matich-redesneuronales.pdf

Merino, J. y Noriega, M.J. (2011). Comunicación neuronal: Sinapsis. Universidad de Cantrabia, Fisiología General, URL: http://ocw.unican.es/ciencias-de-la-salud/fisiologia-general/materiales-de-clase-1/bloque-ii/Tema%208-Bloque%20II-Comunicacion%20Neuronal%20Sinapsis.pdf

Montaño. J.J. y Palmer, A. (1999). ¿Qué son las redes neuronales artificiales? Aplicaciones realizadas en el ámbito de las adicciones. *Adicciones*, URL: http://disi.unal.edu.co/~lctorress/RedNeu/LiRna001.pdf

Montaño, J.J. (2002). Redes neuronales artificiales aplicadas al análisis de datos. (Tesis Doctoral). Facultad de Psicología, Universitat de Les Illes Balears.

Montaño, J.J. (2005). Redes Neuronales Artificiales. Fundamentos, modelos y software. Facultad de Psicología. Universidad de las Islas Baleares.

Morales, P. (2011). Células de Schwann. *La guía de Biología*, URL: http://biologia.laguia2000.com/citologia/celula-de-schwann

Oja, E. (1992). Principal Components, Minor Components, and Lineal Neural Networks. *Neural Networks*, *5,* 927-935.

Pineda, F.J. (1989). Recurrent back-propagation and the dynamical approach to neural computation. *Neural Computation*, *1*, 161-172.

Pitarque, A.; Roy, J.F. y Ruiz, J.C. (1998). Redes neuronales vs modelos estadísticos: Simulaciones sobre tareas de predicción y clasificación. *Psicológica*, *19*, 387-400.

Romero, A. (2010). Nuevos Paradigmas para el análisis estadístico de imágenes tomográficas cerebrales. Editorial de la Univeridad de Granada, URL: http://hera.ugr.es/tesisugr/19486662.pdf

Rosenblatt, F. (1958). The Perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408.

Sanger, T.D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, *2*, 459-473.

Specht, D.F. (1990). Probabilistic neural networks. *Neural Networks*, *3*, 109 -118.

Tezanos, P. (2014). Nódulos de Ranvier: un pequeño espacio que marca la diferencia. *Antroporama, Neurociencia*. URL: http://antroporama.net/nodulos-de-ranvier-un-pequeno-espacio-que-marca-la-diferencia/

Villanueva, Mº.R. (2002). Las redes neuronales artificiales y su importancia como herramienta en la toma de decisiones. Universidad Nacional de San Marcos, Facultad de ciencias matemáticas, URL: http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/Villanueva_EM/enPDF/T_completo.pdf

Von Der Malsburg, C. (1973). Self-organitation of orientation sensitive cells in the striate cortex. *Kybernetik*, *14*, 85-100.

Werbos, P.J. (1990). Backpropagation through time: What it is and how to do it. *Proceedings of the IEEE*, *78*(10), 1550-1560.

Williamson, J.R. (1996). Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps. *Neural Networks*, *9*(5), 881-897.

ANEXO 1

En el siguiente cuadro se presentan todos los hechos históricos ocurridos en el campo de las Redes Neuronales Artificiales.

Año	**Autor**	**Aportación**
1943	McCulloch & Pitts	Primera neurona
1949	Hebb	Regla de aprendizaje hebbiano
1958	Rosenblatt	Perceptrón
1960	Widrow & Hoff	ADELINE y Regla delta
1967	Grossberg	AVALANCHA
1969	Minsky & Papert	Perceptrons (crítica conexionismo)
1973	Von Der Malsburg	Columnas Organizadas
1975	Fukushima	Cognitrón
1977	Kohonen	LVQ y TPM
	Anderson	BSB
1980	Grossberg	ART (Teoría Resonancia Adaptativa)
1982	Hopfield	Redes dinámicas
1983	Fukushima	Neocognitrón
1985	Hinton & Sejnowski	Máquina de Boltzmann
1986	Rumelhart, Williams & Hinton	Algoritmo backpropagation
1988	Minsky & Papert	Segunda crítica conexionismo

ANEXO 2

Una neurona tiene un determinado estado de activación en un momento determinado, A(*t*). Visto de una forma más sencilla el estado de activación de una red neuronal pueden ser dos, reposo y excitado, y cada uno de ellos tiene un valor.

Las conexiones que unen las neuronas tienen un peso asociado, w_{ij}, que es el que hace que la red adquiera el conocimiento. La entrada que recibe cada neurona tiene un valor aditivo, cada entrada es el sumatorio del producto de cada señal individual por el peso de la conexión. En la Figura Anexo 1 se muestra la expresión de la señal de entrada de una neurona *j*.

$$net_j = \sum_i^N w_{ij} y_i$$

Figura Anexo 1: Señal de entrada

Para saber el valor de salida que da una red se aplica la llamada función de activación o de salida. Ésta es una regla que combina las entradas de la neurona con su estado actual para producir un nuevo estado de activación.

La expresión usada de esta función de activación es mostrada en la Figura Anexo 2:

$$a_j(t+1) = f(a_j(t), net_j)$$

Figura Anexo 2: Función de activación

El nuevo estado de activación se encuentra representado por aj (t+1) se consigue a partir de la aplicación de una función f al estado de activación actual de la neurona y su entrada neta. Pero en la mayoría de los casos es el estado actual de la neurona no se tienen en cuenta, con

lo que en vez de tener en cuenta este se tiene en cuenta las características internas de la neurona, como es el umbral de activación (θ_j), que corresponde a la energía mínima necesaria para que una neurona se active (ver Figura Anexo 3).

$$y_j(t+1) = f(net_j - \theta_j) = f\left(\sum_{i=1}^{N} w_{ij} y_i(t) - \theta_j\right)$$

Figura Anexo 3: Función de activación 2

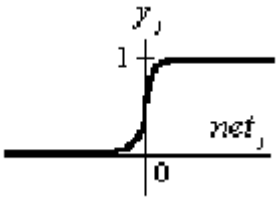
Para simplificar esta expresión, se incluye en la existencia de una neurona ficticia, con un valor de salida equivalente a 1, y que se asocia con la entrada a la neurona *j* a partir de la conexión con el peso de valor $-\theta_j$. Esto se muestra en la Figura Anexo 4.

$$y_j(t+1) = f\left(\sum_{i=1}^{N} w_{ij} y_i(t) - \theta_j \cdot 1\right) = f\left(\sum_{i=0}^{N} w_{ij} y_i(t)\right) = f(net_j)$$

Figura Anexo 4: Función de activación 3

Ahora sabemos cómo se calcula el nuevo estado de activación de la neurona a partir de la aplicación de una función *f* determinada. Esta función puede ser de diferentes tipos en dependiendo del valor de las conexiones.

La función de activación que se suele usar en el Perceptrón Multicapa, es la función sigmoidal (mostrada en la Figura Anexo 5). En esta función los valores de la señales de entrada se definen como un intervalo con un incremento monotónico. La función más común es la sigmoidal, en la que la mayoría de los valores de entrada son cercanos a los valores asintóticos, con lo que hace que la mayoría de los valores de salida se encuentren en los límites representados por los valores [0; 1] ó [-1; 1].



$$f_j(net_j) = \frac{1}{1+e^{-net_j}}$$

Figura Anexo 5: Función de activación sigmoidal

En el funcionamiento del Perceptrón Multicapa se diferencian tres etapas: aprendizaje, funcionamiento y test. Para ello se crean tres subgrupos de muestra (entrenamiento, validación y test) que se aplicarán en cada una de las etapas. El algoritmo backpropagation error que se usa en el funcionamiento del Perceptrón interviene en las fases de aprendizaje y funcionamiento.

Etapas de funcionamiento

Etapa de aprendizaje

La primera etapa, la de aprendizaje, se pasa el grupo de entrenamiento. Hay que tener en cuenta a la hora de crear el grupo de entrenamiento el número de pesos, ya que si este es excesivo en relación al problema aparece el fenómeno de sobreajuste, es decir, el modelo le da demasiada importancia a las peculiaridades de los patrones, lo cual hace que se pierda la habilidad de generalización del aprendizaje.

En esta etapa se modifican los pesos de forma iterativa en función de los valores de este grupo para minimizar el error entre la señal de salida obtenida y la deseada.

La función de error usada para minimizar el error es la que se muestra en la Figura Anexo 6.

$$E_p = \frac{1}{2}\sum_{k=1}^{M}(d_{pk} - y_{pk})^2$$

Figura Anexo 6: Función de error

donde d_{pk} es la salida deseada por una neurona, la neurona *k*, ante un patrón *p*. Con esta expresión podemos conseguir una medida general del error, a partir de la expresión mostrada en la Figura Anexo 7.

$$E = \sum_{p=1}^{P} E_p$$

Figura Anexo 7: Medida general del error

La base matemática de este algoritmo para la modificación de los pesos es la regla del gradiente decreciente o regla delta, creada por Rumelhart, Hinton y Williams en 1960. Teniendo en cuenta que el Ep es en función de todos los pesos de la red, el gradiente del mismo equivale a la derivada parcial de Ep respecto a cada uno de los pesos. Y el gradiente toma la dirección que determina el incremento más del error, por lo tanto la dirección contraria determinará el decremento más rápido (ver Figura Anexo 8).

$$-\sum_{p=1}^{P}\frac{\partial E_p}{\partial w_{ij}}$$

Figura Anexo 8: Dirección del gradiente

Un problema que puede aparecer al utilizar este método, es que el aprendizaje converja en un punto bajo pero sin que este sea el más bajo. Este tipo de puntos se denominan mínimos locales y el punto más bajo de la superficie, se llama mínimo global.

Para que los pesos se modifiquen de forma iterativa se aplica la regla de la cadena a la del gradiente decreciente más una tasa de aprendizaje η. El valor delta que se obtiene está determinado por la suma de los errores que se comenten en cada una de las neuronas.

Además de la tasa de aprendizaje también se añade un factor momento (α), el cual en cuenta la dirección del incremento del gradiente de la conexión anterior. Cada expresión es diferente para una neurona de salida (Figura Anexo 9) que para una neurona oculta (Figura Anexo 10).

$$\Delta v_{jk}(n+1) = \eta\left(\sum_{p=1}^{P} \delta_{pk} b_{pj}\right) + \alpha \Delta v_{jk}(n)$$

Figura Anexo 9:Expresión neurona de salida

$$\Delta w_{ij}(n+1) = \eta\left(\sum_{p=1}^{P} \delta_{pj} x_{pi}\right) + \alpha \Delta w_{ij}(n)$$

Figura Anexo 10: Expresión neurona oculta

La neurona lleva a cabo el aprendizaje, es decir, la actualización de los pesos, tras la presentación de cada patrón de entrenamiento (aprendizaje on line). Es importante que la presentación de los patrones sea aleatoria para no influir en el aprendizaje.

Etapa de funcionamiento

En la etapa de funcionamiento se presentan una serie de patrones de entrada X_p (X_{p1}, X_{p2}, …) que se transmite a través de los pesos wij hacia la capa oculta. Las neuronas de esta capa transforman las señales recibidas a partir de la función de activación, la cual proporciona un

valor de salida que se transmite a la siguiente capa de neuronas, en donde, a través de la aplicación de la misma función se obtiene una señal de salida de la red.

Por lo tanto, una neurona *j* recibe una señal de entrada de la neurona *p*, netpj, (el valor del cual se obtiene con la función vista anteriormente) y esta neurona *j* envía una señal de salida, bpj, a partir de la aplicación de una función sobre su entrada neta.

Etapa test

La etapa test, la cual es la última de todas, sirve para medir de una forma objetiva la eficacia final del sistema que hemos construido. Por lo tanto, se crea el tercer grupo de datos independientes que proporcionará una estimación insesgada del error de generalización.

Consideraciones a tener en cuenta

Una vez visto el funcionamiento se han de tener en cuenta una serie de parámetros que son importantes a tener en cuenta para la creación de una red, pero estos no se pueden conocer a priori sino que se han de establecer una vez visto su funcionamiento, y se seleccionarán aquellos que den el menor error. Estos son

- **Elección de los pesos iniciales:** los pesos de umbral y de conexión inicial se pueden asignar de forma aleatoria. Pero para que el entrenamiento tenga una menor duración se puede seguir una sencilla regla.

  Ésta consiste en la asignación de valores pequeños a los pesos, normalmente comprendidos entre (-0,5; 0,5). De esta forma se consigue que la entrada neta de cada

unidad sea próxima a cero, con lo que el valor que devuelve la función de activación (sigmoidal) es un valor intermedio. De esta forma se puede conocer el intervalo de valores que darán el mínimo error, además de poder conocer los valores extremos, en los cuales el aprendizaje sería nulo.

- **Arquitectura de la red**: esto es para saber el número de neuronas que formarán cada capa. La capa de entrada y de salida han de tener un número de neuronas equivalente al número de variables predictoras presentes. Por ejemplo, si las variables son orientación sexual (heterosexual, homosexual y bisexual) usaremos tres neuronas de entrada y de salida. Cabe destacar que la activación de las neuronas de salida dependerán del tipo de tarea que se esté llevando a cabo.

  En la capa de neuronas ocultas la cantidad de neuronas dependerá de la capacidad de aprendizaje de la red, pero para evitar el problema de sobreajuste siempre se ha de usar la cantidad mínima de neuronas ocultas con las que la red puede funcionar de forma eficaz.

- **Valor de la tasa de aprendizaje y factor momento**: la tasa de aprendizaje es la que controla el tamaño del cambio de los pesos. Se ha de evitar que el ritmo de aprendizaje de la red sea demasiado lento , ya que tendrá lugar una importante disminución en la velocidad de convergencia y aumentará la posibilidad de acabar en un mínimo local, o demasiado rápido, lo que acabaría en la inestabilidad de la función de error y no se llegaría alcanzar el mínimo global. Por lo tanto, para evitar ambos problemas se recomienda usar una tasa de aprendizaje situada entre (0,05 y 0,5).

En cuanto al factor momento, el cual acelera la convergencia de los pesos. El valor que se suele dar a este factor es 1.

- **Función de activación de las neuronas ocultas y de salida**: el requisito de esta función es que ha de ser continua, por lo tanto las funciones básicas que se pueden aplicar son: la función lineal y la sigmoidal.

  En la capa de neuronas ocultas se usará una función sigmoidal para conseguir que la red sea capaz de aprender relaciones complejas. En cambio, en las neuronas de la capa de salida la elección de la función dependerá del tipo de tarea, por ejemplo, para tareas de predicción se usará una función lineal, pero para una tarea de clasificación se recomienda el uso de una función sigmoidal.

ANEXO 3

En este apartado se muestran los valores que se han modificado con los porcentajes de acierto en la clasificación de patrones y el error de la red.

Tabla Anexo 1: Datos de fase de Entrenamiento

Modelo	**Número neuronas ocultas**	**Tasa de aprendizaje**	**Factor Momento**	**RMC Error total**	**Acierto en la clasificación de patrones (%)**
Modelo 1	2	0,1	0,9	0,35	84
Modelo 2	3	0,1	0,9	0,33	86
Modelo 3	4	0,1	0,9	0,31	88
Modelo 4	1	0,1	0,9	0,37	82
Modelo 5	1	0,05	0,9	0,38	82
Modelo 6	2	0,05	0,9	0,35	83
Modelo 7	3	0,05	0,9	0,32	83
Modelo 8	4	0,05	0,9	0,29	89
Modelo 9	1	0,2	0,9	0,37	82
Modelo 10	1	0,05	1	0,39	80
Modelo 11	1	0,05	0,8	0,38	81
Modelo 12	2	0,2	0,9	0,36	79
Modelo 13	2	0,1	1	0,36	83
Modelo 14	2	0,1	0,8	0,35	83

Tabla 2: Datos de fase de Validación

Modelo	**Número neuronas ocultas**	**Tasa de aprendizaje**	**Factor Momento**	**RMC Error total**	**Acierto en la clasificación de patrones (%)**
Modelo 1	2	0,1	0,9	0,405	81
Modelo 2	3	0,1	0,9	0,41	80
Modelo 3	4	0,1	0,9	0,46	71
Modelo 4	1	0,1	0,9	0,41	77
Modelo 5	1	0,05	0,9	0,41	77
Modelo 6	2	0,05	0,9	0,44	74
Modelo 7	3	0,05	0,9	0,45	69
Modelo 8	4	0,05	0,9	0,46	75
Modelo 9	1	0,2	0,9	0,41	77
Modelo 10	1	0,05	1	0,43	73
Modelo 11	1	0,05	0,8	0,41	77
Modelo 12	2	0,2	0,9	0,45	69
Modelo 13	2	0,1	1	0,43	75
Modelo 14	2	0,1	0,8	0,44	74