UNIVERSITAT JAUME I

Grado en Matemática Computacional

Proyecto Final de Grado

Un resultado de aproximación de funciones continuas mediante redes neuronales

Autor:
Andrea Gómez Gelo

Tutores académicos:
Juan José Font Ferrandis
Sergio Macario Vives

Fecha de lectura: Octubre de 2022
Curso académico 2021/2022

Resumen

Este documento recoge el Proyecto de la asignatura *MT-1054/ Trabajo de Final de Grado*, del Grado en Matemática Computacional, cursado en la Universitat Jaume I.

Este trabajo consta de tres capítulos. En el primero de ellos, se introduce el tema a tratar y los objetivos que persigue este proyecto. A continuación, en el segundo capítulo, se presentan las definiciones necesarias junto con el estudio del error en la aproximación de una función continua de una variable mediante redes neuronales activadas usando una función sigmoidal. Para concluir, en el último capítulo se presentan las conclusiones.

Palabras clave

Red neuronal, función sigmoidal, función continua.

Abstract

This document details the Project of the subject *MT-1054 / Final Degree Project*, of the Degree in Computational Mathematics, at the Universitat Jaume I.

It consists of three chapters. In the first one, the topic to be discussed and the objectives pursued by this project are introduced. Next, in the second chapter, some definitions are presented with the study of the error of approximation of a continuous map with one variable using neural networks activated by a sigmoidal map. To conclude, the last chapter presents the conclusions.

Keywords

Neural network, sigmoidal map, continuous map.

Agradecimientos

Me gustaría agradecer a todas y cada una de las personas que han formado parte de esta gran aventura y han ayudado a que hoy esté aquí. Han pasado 4 años desde que comencé el grado de Matemática Computacional, y aunque ha sido una montaña rusa, con el tiempo he aprendido a disfrutar de las bajadas gracias a todas las cosas nuevas que han ido apareciendo.

Gracias a mis compañeros de clase, porque desde que nos conocimos con 18 años hemos sido una pequeña gran familia, hemos compartido muchas horas, muchos momentos y sobre todo muchas risas, que de eso trata la vida. Me gustaría destacar a esas personas que pasaron de ser compañeros de clase a ser amigos. Parece que fue ayer cuando compartíamos horas de estudio, comidas en la pecera y mucha frustración. Pero también creamos muchos buenos momentos que nos hicieron disfrutar en estos 4 años. Aunque cada uno haya tomado un camino distinto, todos habéis marcado esta experiencia.

Por otra parte, quiero agradecer a mis padres, por la confianza que siempre han tenido en mi, aún cuando ni yo creía que fuera capaz de lograrlo. Ha sido una suerte saber que cuando las cosas no iban bien, siempre os iba a tener ahí para animarme. Por haberme dado siempre la libertad de hacer lo que me apasionara y celebrar todos mis triunfos.

A mi prima Lourdes, por convertirse en mi primer referente cuando yo era pequeña. Por haber sido un ejemplo de superación y demostrarme que lo importante de la vida es ser tu misma. Gracias por dejarme compartir tantas horas contigo y dejarme ver que la vida con intensidad es vida igual.

A mi prima Silvia, porque con ella descubrí que no era la única amante de las matemáticas en la familia. Gracias por haber cursado esta carrera antes que yo y haberme dado los mejores consejos. Por haber estado siempre ahí, escuchando todos mis dramas y sacando una de mis mejores versiones.

Por último, me gustaría agradecer a mis tutores, Juanjo y Sergio, por ayudarme en la realización de este trabajo. Por su paciencia, dedicación y ayuda. Ha sido una gran experiencia.

Índice general

Índice de figuras

Capítulo 1

Introducción

1.1. Contexto y motivación del proyecto

En este documento se presenta el Trabajo Final de Grado, donde se expondrá un estudio de una de las diversas capacidades de las redes neuronales en el ámbito de las matemáticas como es la aproximación de una función continua de una variable mediante una red neuronal con una capa oculta.

En la actualidad, vivimos en un mundo que está en constante desarrollo, sobre todo en el mundo tecnológico. Dentro de este ámbito, se han ido popularizando términos relacionados con la inteligencia artificial y el aprendizaje automático. Dentro de estos, podemos encontrar un nuevo concepto conocido como redes neuronales.

Las redes neuronales están jugando un papel importante en diferentes campos y áreas del conocimiento, incluyendo la medicina como parte de la inteligencia artificial.

Una red neuronal presenta una estructura similar a la siguiente:

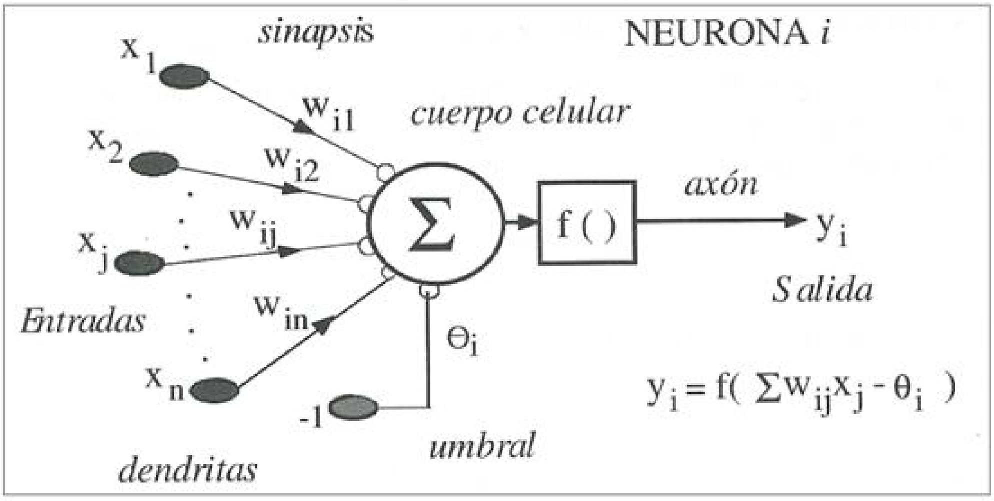

Figura 1.1: Estructura red neuronal

Dicha estructura está formada principalmente por un conjunto de nodos que simulan a las neuronas y que están conectadas entre sí. El funcionamiento de estas comienza cuando reciben una información de entrada que procesan conjuntamente para después generar una respuesta. Para entender su funcionamiento, solo tenemos que fijarnos en un cerebro humano, en el que millones de neuronas forman una red y mediante unas conexiones o sinapsis consiguen comunicarse las unas con las otras para realizar una acción.

El objetivo de las redes neuronales es imitar los procesos mediante los cuales los seres humanos toman decisiones y llevarlo a un computador. Además, cuentan con la funcionalidad de que también pueden formarse a sí mismas, es decir, entrenar con conjuntos de entrada para aprender. Esto es un avance, ya que no necesitan ser programadas explícitamente, sobresaliendo en áreas donde es difícil la programación convencional para extraer soluciones.

Matemáticamente, estas se pueden expresar como una terna de la forma $N =< D, f_i, A >$ donde:

1. D representa un digrafo formado por unos vértices que simulan las neuronas conectadas entre sí mediante arcos que muestran las conexiones sinápticas con unas etiquetas llamadas pesos que indican la intensidad de estas conexiones.

2. A representa un conjunto que contiene los elementos de entrada de las neuronas nombrado como Activación.

3. f_i es un conjunto de funciones llamadas de transferencia

Este concepto lleva presente desde la década de los 40, cuando McCulloch y Pitts dieron la primera definición en su artículo [7], donde una red neuronal se definía como una máquina binaria con varias entradas y salidas.

Esto dio paso a que otros científicos se interesaran y decidieran dedicarse a su estudio. Minsky y Pappert, Rosenblatt, Cohen y Gros, ... son algunos ejemplos de científicos que aportaron información sobre este campo, ya fuera mediante algoritmos, artículos de investigación o desarrollos.

En el contexto en que nos movemos, el tipo de red neuronal que utilizaremos es aquella con una capa oculta. Se define como una función, N, que transforma un elemento $x \in \mathbb{R}$ en un número real de la siguiente forma:

$$N(x) = c_0 + \sum_{i=1}^{n} c_i \cdot \phi(w_i \cdot x + \theta_i)$$

donde los coeficientes $w_i, c_i, \theta_i \in \mathbb{R}$ son los pesos y la función $\phi : \mathbb{R} \to \mathbb{R}$ es la denominada *función de activación.*

Cybenko nos muestra en su estudio, [2], los primeros resultados sobre la aproximación de funciones usando redes neuronales activadas mediante funciones sigmoidales. Otros ejemplos de investigaciones interesantes sobre este tipo de función son las que realizan Hong y Hahm en [4], sobre la aproximación de la red neuronal a una función continua en $\mathbb{R}$ y [5], donde se sugiere una aproximación constructiva por las redes neuronales con una función sigmoidal utilizando el método de convolución.

Kalman y Kwasny enfatizan en su estudio, [6], la importancia de las redes neuronales activadas mediante una función tangente hiperbólica, debido a las aplicaciones en implementación de hardware de propagación.

Por último, en el estudio de Medvedeva, [8], se sugiere un resultado de densidad por redes neuronales con una función sigmoidal usando el teorema de Taylor.

La motivación principal de este proyecto es, por tanto, conocer y comprender el funcionamiento de las redes neuronales en el contexto de la aproximación de funciones. Este tema es muy interesante y atractivo, ya que está muy presente en nuestro día a día debido a sus diversas aplicaciones y a su rápido desarrollo en ámbitos tecnológicos.

En el segundo capítulo, procedemos al estudio de los resultados teóricos mediante una serie de técnicas básicas de Análisis Matemático aprendidas durante el grado. Además, también se ha incluido un apartado que contiene los conceptos previos que se utilizarán en los resultados teóricos posteriores.

El resultado principal se muestra en el Teorema 1. Para concluir con el capítulo presentaremos un ejemplo numérico, implementado en Matlab, que permitirá mostrar tanto de manera gráfica como numérica el resultado principal de la memoria.

Finalmente, en el tercer capítulo se presentan las conclusiones principales del proyecto, así como una reflexión acerca de los ámbitos personal y formativo.

Capítulo 2

Aproximación mediante redes neuronales

2.1. Resultados teóricos

2.1.1. Conceptos previos

En esta parte se van a presentar una serie de definiciones necesarias para el estudio que se va a realizar. En todo momento, se considerará que estamos trabajando con una red neuronal como la definida en la Introducción.

Definición 1 *Una función sigmoidal es una aplicación*

$$\sigma : \mathbb{R} \longrightarrow \mathbb{R}$$

que cumple $\lim_{t \to -\infty} \sigma(t) = k_1$ *y* $\lim_{t \to +\infty} \sigma(t) = k_2, \quad k_1, k_2 \in \mathbb{R}$

Las funciones de activación sigmoidales más habituales son las siguientes:

Función de Heaviside o escalón unitario

$$\sigma(x) = \begin{cases} 1 & \text{si} \quad x > 0 \\ \\ 0 & \text{si} \quad x \leq 0 \end{cases}$$

Función lineal a trozos o rampa unidad:

$$\sigma(x) = \begin{cases} 1 & \text{si} \quad x \geq 1 \\ \\ x & \text{si} \quad 0 \leq x < 1 \\ \\ 0 & \text{si} \quad x < 0 \end{cases}$$

Función logística:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Esta última, en particular, se lleva utilizando desde el siglo XIX para modelar el crecimiento de poblaciones. Hoy en día es muy utilizada en ámbitos como la Estadística, la Física, la Biología, etc.

Definición 2 *Una función $f : X \longrightarrow \mathbb{R}$ se dice que está **acotada** si, existe un número real $M > 0$ para todo $x \in X$ tal que $|f(x)| \leq M$.*

Definición 3 *Una función $f : \mathbb{R} \longrightarrow \mathbb{R}$ se dice que es **continua** en un punto $x_0 \in \mathbb{R}$ si, para cada $\epsilon > 0$, se puede encontrar $\delta > 0$ tal que, para todo $x \in \mathbb{R}$ con $|x - x_0| < \delta$, se verifica que $|f(x) - f(x_0)| < \epsilon$. O también, si para cada sucesión $(x_n)_{n=1}^{\infty} \subset \mathbb{R}$ con $\lim_{n\to\infty} x_n = x_0$, se tiene que $\lim_{n\to\infty} f(x_n) = f(x_0)$. Tendremos **continuidad uniforme** cuando δ no dependa de x_0.*

En particular, una función continua $f : [a, b] \to \mathbb{R}$ es uniformemente continua.

Definición 4 *Sea $\mathbb{V}$ un espacio vectorial. Una función $|| \cdot || : \mathbb{V} \longrightarrow \mathbb{R}^+$ es una norma en $\mathbb{V}$ si satisface las siguientes propiedades:*

$$\begin{aligned} &(i)\ x \neq 0 \Leftrightarrow ||x|| > 0. \\ &(ii)\ ||\alpha x|| = |\alpha| ||x||, \forall \alpha \in \mathbb{R}. \\ &(iii)\ ||x + y|| \leq ||x|| + ||y||, \forall x, y \in \mathbb{V}. \end{aligned} \tag{2.1}$$

Definición 5 *En el espacio de las funciones continuas de [a,b] en R definimos la norma* $||f||_\infty$ *como el valor máximo de* f *en el intervalo [a,b]:*

$$||f||_\infty = max_{x\in[a,b]}(|f(x)|) \tag{2.2}$$

Definición 6 *Una función* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *se dice que es impar si, para todo* $x \in \mathbb{R}$, *se cumple que* $f(-x) = -f(x)$.

Definición 7 *Una función* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *se dice que es:*

- *estrictamente creciente cuando* $x < y$ *implica* $f(x) < f(y)$
- *estrictamente decreciente cuando* $x < y$ *implica* $f(x) > f(y)$
- *estrictamente monótona cuando es estrictamente creciente o estrictamente decreciente.*

Toda función $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ *estrictamente monótona es inyectiva y, por tanto, tiene inversa* $\phi^{-1} : \phi(\mathbb{R}) \longrightarrow \mathbb{R}$.

Definición 8 *Dada una función continua* $f : [a,b] \to \mathbb{R}$, *definimos su módulo de continuidad como:*

$$\omega(f,\delta) = \sup\ \{|f(s) - f(t)| : |s-t| \leq \delta;\ s,t \in [a,b]\}$$

donde $\delta > 0$. *Tenemos que* $lim_{\delta\to 0}\ \omega(f,\delta) = 0$ *y, para cualquier número real* $\lambda \geq 0$,

$$\omega(f,\delta\lambda) \leq (\lambda+1)\omega(f,\delta)$$

Definición 9 *Dada una función* $f : \mathbb{R} \longrightarrow \mathbb{R}$, *se dice que es lipschitziana de tipo* α $(0 < \alpha \leq 1)$ *si existe una constante* $M > 0$ *tal que* $\omega(f,\delta) \leq M \cdot \delta^\alpha$. *Escribiremos* $f \in Lip_M(\alpha)$.

2.1.2. Resultado principal

Teorema 1 *Sea ϕ una función acotada, estrictamente monótona e impar definida en $\mathbb{R}$. Sean $f \in C[a,b]$ y $n \in \mathbb{N}$. Entonces existe una red neuronal con una capa oculta definida como:*

$$N_n(x) = c_0 + \sum_{i=1}^{n} c_i \phi(w_i x + \theta_i)$$

que cumple que:

$$||N_n - f||_\infty = \sup_{x\in[a,b]} |N_n(x) - f(x)| \leq \frac{5}{2}\omega\left(f, \frac{b-a}{n}\right)$$

donde los parámetros c_i y w_i están definidos como:

$$x_i = a + i\frac{b-a}{n}, \quad x_0 = a, \quad m = \sup_{x\in\mathbb{R}} |\phi(x)|, \quad c_i = \frac{1}{2m}(f(x_i) - f(x_{i-1}))$$

$$d_n = \phi^{-1}\left(m - \frac{m}{2n}\right), \quad w_i = \frac{2d_n n}{b-a}$$
$$\theta_i = -\frac{d_n n}{b-a}\left(2a + (2i-1)\frac{b-a}{n}\right)$$

,

$$c_0 = f(a) - \sum_{i=1}^{n} \phi(w_i a + \theta_i)$$

Demostración

El objetivo es demostrar que para todo $x \in [a,b]$ se cumple que:

$$|N_n(x) - f(x)| \leq \frac{5}{2}\omega\left(f, \frac{b-a}{n}\right)$$

Comenzaremos la demostración dividiendo el intervalo $[a,b]$ en n segmentos iguales, cuya longitud es $\dfrac{b-a}{n}$ donde $a = x_0 < x_1 < ... < x_n = b$, siendo $x_i = a + \dfrac{i(b-a)}{n}$.

Sea $m = \sup_{x\in\mathbb{R}} |\phi(x)|$. Para cada $1 \leq i \leq n$, se definen los siguientes términos:

$$c_i = \frac{1}{2m}(f(x_i) - f(x_{i-1})),\ d_n = \phi^{-1}\left(m - \frac{m}{2n}\right)$$

$$\theta_i = -\frac{d_n n}{b-a}(x_i + x_{i-1}),\ w_i = \frac{2d_n n}{b-a}$$

Como la función ϕ es estrictamente monótona, entonces, ϕ es una función inyectiva. Esto, implica, que la función ϕ tiene inversa en $\phi(\mathbb{R})$.

Sea $c_0 = f(a) - \sum_{i=1}^{n} c_i\phi(w_i a + \theta_i)$.

Ahora, construimos una red neuronal con una única capa oculta definida como:

$$N_n(x) = c_0 + \sum_{i=1}^{n} c_i\phi(w_i x + \theta_i)$$

Notamos que con la elección de c_0 garantizamos que $N_n(a) = f(a)$ y, además, se cumple que $-m \leq \phi(x) \leq m$, $x \in \mathbb{R}$.

Fijamos $n \in \mathbb{N}$. Para cualquier $x \in [a,b]$, existe un $j \in \mathbb{N}$ con $0 < j \leq n$, tal que $x \in [x_{j-1}, x_j]$. Por una parte:

$$|N_n(x) - f(x)| \leq |N_n(x) - f(x_{j-1})| + |f(x) - f(x_{j-1})|,$$

y probaremos que cada término puede acotarse convenientemente para obtener el resultado.

Seguidamente, a partir de la elección de c_0, podemos escribir:

$$N_n(x) = c_0 + \sum_{i=1}^{n} c_i\phi(w_i x + \theta_i) = f(a) - \sum_{i=1}^{n} c_i\phi(w_i a + \theta_i) + \sum_{i=1}^{n} c_i\phi(w_i x + \theta_i)$$

$$= f(a) + \sum_{i=1}^{n} \frac{1}{2m}(f(x_i) - f(x_{i-1}))(\phi(w_i x + \theta_i) - \phi(w_i a + \theta_i)).$$

Llamamos $F_i(x) = \phi(w_i x + \theta_i)$ y, entonces:

$$N_n(x) = f(a) + \sum_{i=1}^{n} \frac{1}{2m}(f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a))$$

$$= f(a) + \sum_{i=1}^{j-1} \frac{1}{2m}(f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a)) + \frac{1}{2m}(f(x_j) - f(x_{j-1}))(F_j(x) - F_j(a))$$

$$+ \sum_{i=j+1}^{n} \frac{1}{2m}(f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a)) = f(a) + \Sigma_1 + \Sigma_2 + \Sigma_3$$

A continuación, vamos a proceder a estudiar los términos Σ_1, Σ_2 y Σ_3 por separado.

El término Σ_1 puede escribirse como:

$$\Sigma_1 = \frac{1}{2m} \sum_{i=1}^{j-1}(f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a) - 2m) + (f(x_{j-1}) - f(a))$$

puesto que

$$\frac{1}{2m} \sum_{i=1}^{j-1} 2m(f(x_i) - f(x_{i-1})) = f(x_1) - f(x_0) + f(x_2) - f(x_1) + ... + f(x_{j-1}) - f(x_{j-2})$$

$$= f(x_{j-1}) - f(a)$$

de donde $f(a) + \Sigma_1 = \sum_{i=1}^{j-1}(f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a) - 2m) + f(x_{j-1})$.

Por otra parte, para el término Σ_2 resulta que:

$$|\Sigma_2| = \frac{1}{2m}|f(x_j) - f(x_{j-1})||F_j(x) - F_j(a)|$$

Por un lado tenemos que $|f(x_j) - f(x_{j-1})| \leq \omega\left(f, \frac{b-a}{n}\right)$, mientras que por otro lado tenemos,

$|F_j(x) - F_j(a)| = |\phi(w_j \cdot x + \theta_j) - \phi(w_j \cdot a + \theta_j)| \leq |\phi(w_j \cdot x + \theta_j)| + |\phi(w_j \cdot a + \theta_j)| \leq m + m = 2m$

.

Por lo tanto,

$$|\Sigma_2| = \frac{1}{2m}|f(x_j) - f(x_{j-1})||F_j(x) - F_j(a)| \leq \frac{1}{2m}\omega\left(f, \frac{b-a}{n}\right)2m = \omega\left(f, \frac{b-a}{n}\right).$$

Finalmente, para el término Σ_3, se observa que, para cada $i > j$ tenemos que $x \leq x_j \leq x_{i-1}$.

Debido a que ϕ es una función estrictamente monótona, resulta que F_i también lo es y se tiene que:

$$0 < F_i(x) - F_i(a) \leq F_i(x_j) - F_i(a) \leq F_i(x_{i-1}) - F_i(a) \leq F_i(x_{i-1}) + m = \phi(w_i x_{i-1} + \theta_i) + m.$$

Ahora observamos que:

$$w_i x_{i-1} + \theta_i = \frac{2d_n n}{b-a}x_{i-1} - \frac{d_n n}{b-a}x_i - \frac{d_n n}{b-a}x_{i-1} = \frac{d_n n}{b-a}(x_{i-1} - x_i) = -\frac{b-a}{n} \cdot \frac{d_n n}{b-a} = -d_n.$$

Si volvemos a la expresión anterior obtenemos:

$$0 < F_i(x) - F_i(a) \leq F_i(x_j) - F_i(a) \leq F_i(x_{i-1}) - F_i(a) \leq F_i(x_{i-1}) + m = \phi(w_i x_{i-1} + \theta_i) + m$$

$$= \phi(-d_n) + m = -\phi(d_n) + m = m - \phi\left(\phi^{-1}\left(m - \frac{m}{2n}\right)\right) = m - \left(m - \frac{m}{2n}\right) = \frac{m}{2n}$$

por lo que:

$$|\Sigma_3| \leq \sum_{i=j+1}^{n} \left|\frac{1}{2m}(f(x_i) - f(x_{i-1}))\right| (F_i(x) - F_i(a))$$

Por un lado tenemos que $|f(x_i) - f(x_{i-1})| \leq \omega\left(f, \dfrac{b-a}{n}\right)$ mientras que por otro lado tenemos que $F_i(x) - F_i(a) \leq \dfrac{m}{2n}$.

Por lo tanto:

$$|\Sigma_3| \leq \sum_{i=j+1}^{n} \left| \frac{1}{2m}(f(x_i) - f(x_{i-1})) \right| (F_i(x) - F_i(a)) \leq \sum_{i=j+1}^{n} \frac{1}{2m} \omega\left(f, \frac{b-a}{n}\right) \frac{m}{2n}$$
$$\leq \frac{1}{4n} \sum_{i=j+1}^{n} \omega\left(f, \frac{b-a}{n}\right) = \frac{n-j}{4n} \omega\left(f, \frac{b-a}{n}\right).$$

Para cada $1 \leq i < j$, tenemos que $x_i \leq x_{j-1} \leq x$. Entonces:

$2m \geq F_i(x) - F_i(a) > F_i(x_i) - F_i(x_{i-1}) = \phi(w_i x_i + \theta_i) - \phi(w_i x_{i-1} + \theta_i)$

$= \phi\left(\frac{2d_n n}{b-a} \cdot \left(a + i\frac{b-a}{n}\right) - \frac{2d_n n}{b-a}\left(2a + (2i-1)\frac{b-a}{n}\right)\right)$

$- \phi\left(\frac{d_n n}{b-a} \cdot \left(a + (i-1)\frac{b-a}{n}\right) - \frac{d_n n}{b-a}\left(2a + (2i-1)\frac{b-a}{n}\right)\right) = \phi(d_n) - \phi(-d_n)$

$= \phi(d_n) + \phi(d_n) = 2\phi(d_n) = 2\phi\left(\phi^{-1}\left(m - \frac{m}{2n}\right)\right) = 2m - \frac{m}{n}.$

Esto implica que:

$$|F_i(x) - F_i(a) - 2m| \leq \frac{m}{n}$$

Hasta ahora tenemos la red neuronal definida como:

$$N_n(x) = f(a) + \Sigma_1 + \Sigma_2 + \Sigma_3$$

Si sustituimos Σ_1

$$N_n(x) = f(a) + \sum_{i=1}^{j-1} \frac{1}{2m}(f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a) - 2m) + (f(x_{j-1}) - f(a)) + \Sigma_2 + \Sigma_3.$$

Entonces obtenemos

$$N_n(x) - f(x_{j-1}) = \frac{1}{2m} \sum_{i=1}^{j-1} (f(x_i) - f(x_{i-1}))(F_i(x) - F_i(a) - 2m) + \Sigma_2 + \Sigma_3$$

Consecuentemente:

$$
\begin{aligned}
|N_n(x) - f(x_{j-1})| &\leq \frac{1}{2m}\omega\left(f, \frac{b-a}{n}\right)\sum_{i=1}^{j-1}|F_i(x) - F_i(a) - 2m| + |\Sigma_2| + |\Sigma_3| \\
&\leq \omega\left(f, \frac{b-a}{n}\right)\left(\frac{j-1}{2n} + \frac{n-j}{4n} + 1\right) \\
&\leq \frac{3}{2}\omega\left(f, \frac{b-a}{n}\right)
\end{aligned}
$$

Puesto que $x_{j-1} \leq x < x_j$ se cumple:

$$
|f(x) - f(x_{j-1})| \leq \omega\left(f, \frac{b-a}{n}\right),
$$

deducimos que:

$$
\begin{aligned}
|N_n(x) - f(x)| &\leq |N_n(x) - f(x_{j-1})| + |f(x) - f(x_{j-1})| \\
&\leq \left(1 + \frac{3}{2}\right)\omega\left(f, \frac{b-a}{n}\right) = \frac{5}{2}\omega\left(f, \frac{b-a}{n}\right),
\end{aligned}
$$

concluyendo así la demostración. □

Teorema 2 *Sea ϕ una función acotada, estrictamente monótona e impar definida en $\mathbb{R}$. Sea $f \in Lip_M(\alpha)$ con $0 < \alpha \leq 1$ y $n \in \mathbb{N}$. Entonces existe una red neuronal con una capa oculta $N_n : [a,b] \to \mathbb{R}$ definida como:*

$$
N_n(x) = c_0 + \sum_{i=1}^{n} c_i\phi(w_i x + \theta_i)
$$

donde los parámetros c_i y w_i son los definidos en el Teorema 1 tal que:

$$
\sup_{x\in[a,b]} |N_n(x) - f(x)| \leq \frac{5M}{2}\left(\frac{b-a}{n}\right)^{\alpha}
$$

2.2. Ejemplo numérico

En este apartado proporcionamos un ejemplo númerico para ilustrar nuestro resultado teórico usando una función de activación sigmoidal. Para ello, vamos a aproximar una función trigonométrica, implementando en Matlab las técnicas utilizadas en la sección precedente. Consideramos:

$$f(x) = sin(x) \text{ con } x \in [0, \pi]$$

como función a aproximar y la función ϕ como función de activación dada por:

$$\phi(x) = \frac{2}{\pi} arctan(x) \text{ con } x \in \mathbb{R}$$

que cumple los requisitos de ser estrictamente monótona, acotada e impar, además de ser sigmoidal.

La red neuronal que vamos a utilizar es la que aparece en el Teorema 1, es decir:

$$N_n(x) = c_0 + \sum_{i=1}^{n} c_i \phi \Bigg(w_i \cdot x + \theta_i \Bigg) \tag{2.3}$$

donde

$$\begin{aligned}
c_i =& \frac{1}{2} \left(sin \left(\frac{i\pi}{n} \right) - sin(\frac{(i-1)\pi}{n}) \right). \\
\theta_i =& -(2i-1) tan \left(\frac{\pi}{2} \left(1 - \frac{1}{2n} \right) \right). \\
w_i =& \frac{2n}{\pi} tan \left(\frac{\pi}{2} \left(1 - \frac{1}{2n} \right) \right). \\
c_0 =& -\sum_{i=1}^{n} \frac{1}{2} \left(sin \left(\frac{i\pi}{n} \right) - sin \left(\frac{(i-1)\pi}{n} \right) \right) \cdot \frac{2}{\pi} arctan \left((1-2i) tan \left(\frac{\pi}{2} \left(1 - \frac{1}{2n} \right) \right) \right)
\end{aligned} \tag{2.4}$$

Los siguientes bloques de código muestran la definición del coeficiente θ_i necesario para la función de activación y la del coeficiente c_i necesario para el cálculo de la red neuronal. Los coeficientes c_0 y w_i son fijos y se calcularán en el programa principal.

```
% Devuelve el valor de \theta_i segun la posicion del indice
function theta_i =thetai(indice ,n)
    theta_i = -(2*indice - 1)*tan((pi/2)*(1-(1/(2*n))));
end
```

```
% Devuelve el valor de c_i segun la posicion del indice
function c_i = ci(indice ,n)
    c_i = 0.5*(sin(indice*pi/n) - sin((indice - 1)*pi/n));
end
```

Estas dos funciones se utilizan para calcular el valor de un punto en $f(x)$ y en $\phi(x)$:

```
% Devuelve el valor de la funcion de activacion
function z = fi(x)
    z = (2/pi)*atan(x);
end
```

```
%Devuelve el valor de f(x)
function y = f(x)
    y = sin(x);
end
```

El siguiente código devuelve una aproximación al valor de la norma $||N_n - f||_\infty$ calculando el máximo de un vector de diferencias que se obtiene en el código principal:

```
function max = max_error(vector , n)
    max = 0;
    for i = 1:n
        if vector(i)>max
            max = vector(i);
        end
    end
end
```

Por último, el siguiente cuadro muestra el código principal que permite realizar la aproximación y mostrar gráficamente los resulados:

```
% VARIABLES FIJAS
res;
n = 5;
c0;
x = linspace(0,pi);
f_x = sin(x);
w_i = (2*n/pi)*tan((pi/2)*(1-(1/(2*n))));

%Calculo de c0
res = 0;
for i = 1:n
    res = res + 0.5*((sin((i*pi)/n)) - sin((i - 1)*pi/n))
    *atan((1-2*i)*tan((pi/2)*(1-(1/(2*n)))));
end
c0 = -res;

%Construccion de la red neuronal
N_x = c0;
for i=1:n
    N_x = N_x + ci(i,n)*fi(w_i*x + thetai(i,n));
end

% Calculo error
k = 0;
erroraprox = [];

    for y = linspace(0,pi,200)
        k = k+1;
        erroraprox(k) = abs(N_x(y,n) - f(y));
    end

% Resultados
disp(max_error(erroraprox,200));
plot(x,f_x,'r',x,N_x,'k');
legend("f(x)","Red neuronal")
xlim([0 3.25])
ylim([0 1])
```

A continuación, las siguientes gráficas, obtenidas con el código anterior, ilustran como mejora la aproximación al incrementar el valor de n, concretamente para $n = 5$, $n = 10$, $n = 50$, $n = 100$ y $n = 500$.

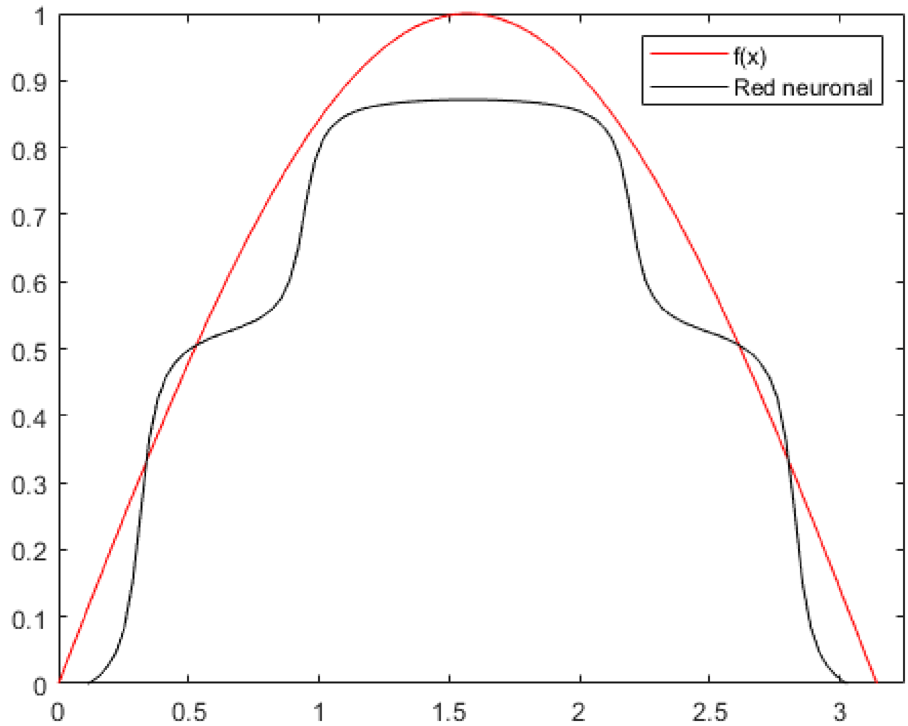

Figura 2.1: Gráfica de la aproximación a $f(x)$ para $n = 5$

En esta representación, podemos observar como la gráfica de la red neuronal no es todavía una buena aproximación de la gráfica de $f(x) = sen(x)$.

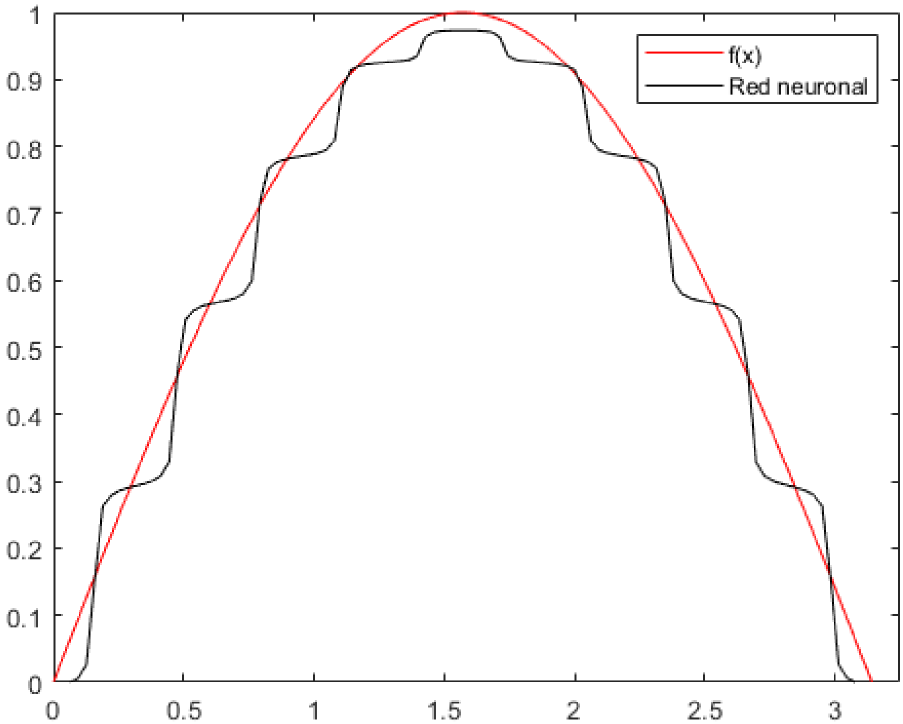

Figura 2.2: Gráfica de la aproximación a $f(x)$ para $n = 10$

En esta representación, podemos observar como al incrementar un poco el valor de n, la gráfica de la red neuronal se aproxima un poco más a la representación de la función, pero sigue presentando bastante error.

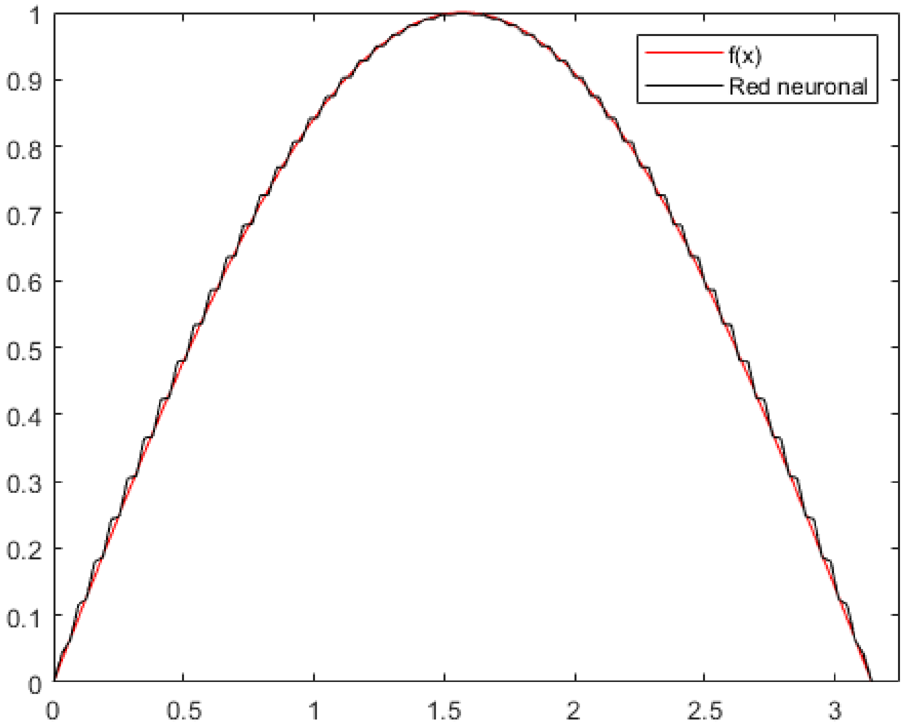

Figura 2.3: Gráfica de la aproximación a $f(x)$ para $n = 50$

En esta figura, ya se observa cómo el incremento del valor de n respecto a la gráfica anterior, mejora la aproximación que buscamos.

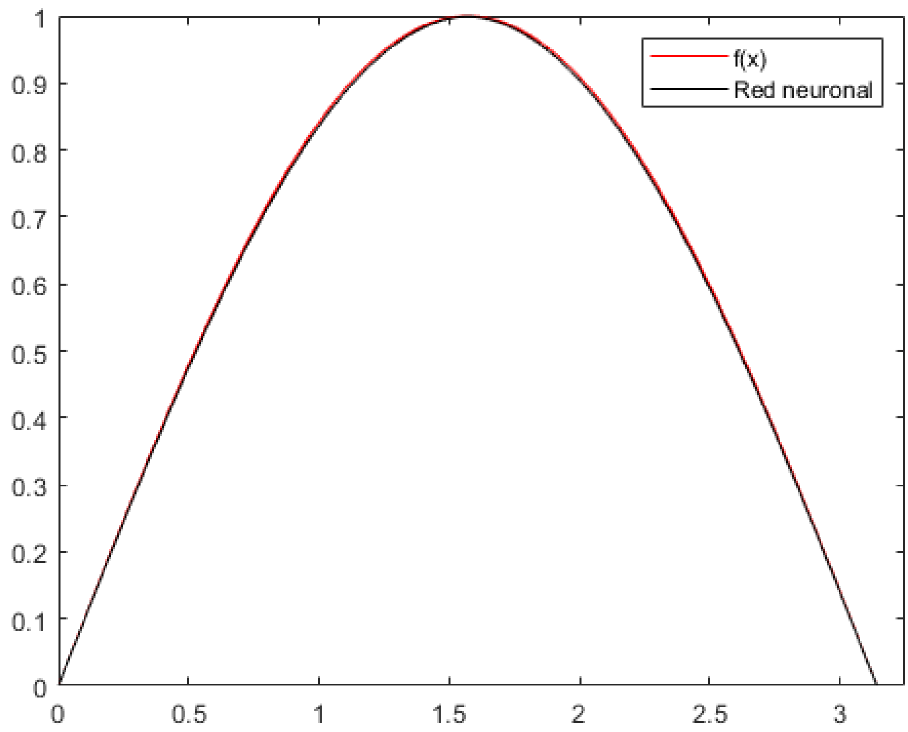

Figura 2.4: Gráfica de la aproximación a $f(x)$ para $n = 100$

En esta gráfica, incrementando el valor de n a 100, vemos como la representación de la red neuronal superpone prácticamente a la representación de la función $f(x) = sen(x)$.

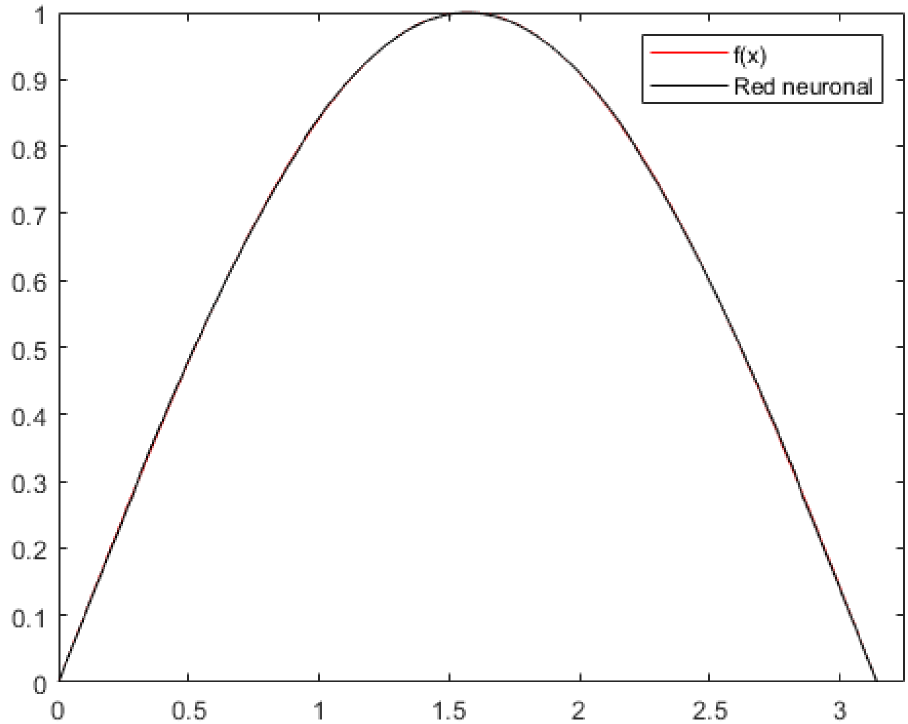

Figura 2.5: Gráfica de la aproximación a $f(x)$ para n = 500

Finalmente, en esta gráfica podemos apreciar como la aproximación mediante la red neuronal es casi perfecta, puesto que su gráfica coincide con la de la función. Esto implicaria que el error de esta aproximación sería un valor muy cercano a 0.

La siguiente tabla de resultados muestra numéricamente lo que hemos podido observar con las gráficas anteriores, es decir, el error máximo entre la red neuronal $N_n(x)$ y la función $f(x) = sin(x)$ va decreciendo a medida que aumentamos el valor de n. Este error máximo es la norma $sup_{x\in[0,\pi]} \; ||N_n(x) - sin(x)||$.

Del Teorema 2 se deduce que $sup_{x\in[0,\pi]} \; |N_n(x) - sin(x)| \leq \dfrac{5\pi}{2n}$:

n	Error máximo	$\frac{5\pi}{2n}$
5	0.1629	1.5708
10	0.0869	0.7854
50	0.0215	0.1571
100	0.0127	0.0785
500	0.0029	0.0157

Cuadro 2.1: Errores máximos para los valores de n

Capítulo 3

Conclusiones

Para algunos conocidas, para otros desconocidas, las redes neuronales son una herramienta muy potente que tiene aplicaciones en diversos campos, incluido las matemáticas, donde son de gran utilidad para la aproximación de funciones. Sin embargo, no siempre es fácil entender cómo se comportan en su capa interna (la de la función de activación) para conseguir esta aproximación.

En este documento se trabaja en uno de los contextos más habituales dentro de este campo: el de las funciones continuas definidas en un intervalo [a,b] y el de las redes neuronales activadas mediante funciones sigmoidales. De hecho, profundizamos con todo detalle en el proceso mediante el cual estas redes neuronales aproximan estas funciones continuas de tal forma que queda patente cuál es el papel que juegan las funciones sigmoidales de activación. Finalmente, un ejemplo numérico corrobora los resultados teóricos.

3.1. Ámbito formativo

En primer lugar, este proyecto me ha servido para poner en práctica muchos de los conocimientos adquiridos durante el Grado de Matemática Computacional cursado en la Universitat Jaume I.

Este trabajo me ha permitido combinar tanto la rama matemática como la rama informática. En el ámbito matemático, he podido desarrollar técnicas de análisis que he ido adquiriendo durante la carrera, empleadas en el estudio de los teoremas. Por otra parte, en el ámbito de la informática se me ha permitido ampliar mis conocimientos acerca de las redes neuronales, además de utilizar estas en programas informáticos.

Además, también he podido ampliar mis conocimientos en la utilización de diversas herramientas como son LaTex o MatLab, utilizadas durante la carrera.

Para concluir, también ha sido muy útil la ayuda recibida por parte de mis tutores, ya que han estado ahí en todo momento para ayudarme y guiarme.

3.2. Ámbito personal

En el ámbito personal, este trabajo me ha ayudado a darme cuenta de que con paciencia se pueden conseguir las cosas. Gracias a él, he podido recordar muchos momentos a lo largo de la carrera, llenos de frustración, cuando las cosas no iban como querías. Pero, poco a poco, y con ganas, se ha ido resolviendo todo.

Empecé esta aventura hace 4 años como una niña amante de las matemáticas que no sabia lo que le esperaba, y hoy se puede decir que sale, aún una niña, pero un poco más matemática.

Bibliografía

[1] F. Cao, T. Xie and Z. Xu, *The estimate for approximation error of neural networks: A constructive approach.* Neurocomputing, **71**, (2008), 626-630.

[2] G. Cybenko, *Approximation by Superpositions of a Sigmoidal Function.* Mathematics of Control, Signals, and Systems, **2**, (1989), 303-314.

[3] B. I. Hong and N. Hahm, *A Note on Neural Network Approximation with a Sigmoidal Function.* Applied Mathematical Sciences, **10**, (2016), no. 42, 2075 - 2085.

[4] B. I. Hong and N. Hahm, *Approximation order to a function in C (R) by superposition of a sigmoidal function.* Applied Mathematics Letters, **15**, (2002), 591-597.

[5] B. I. Hong and N. Hahm, *Constructive approximation by neural networks with positive integer weights.* Korean Journal of Mathematics, **23**, (2015), no. 3, 327-336.

[6] B. L. Kalman and S. C. Kwasny,, *Why tanh: choosing a sigmoidal function.* IJCNN International Joint Conference on Neural Networks **4**, (1992), 578- 581.

[7] W. S. Mcculloch and W. Pitts, *A logical calculus of the ideas immanent in nervous activity.* Bulletin of Mathematical Biophysics, (1943), no 5, 115-133.

[8] M. V. Medvedeva, *On sigmoidal functions.* Moscow University Mathematics Bulletin, **53**, (1998), no. 1, 16-19.