



GRADO EN MATEMÁTICA COMPUTACIONAL

PROYECTO FINAL DE GRADO

Estudio del error en la aproximación de funciones mediante redes neuronales

Autor:
Estefanía GONZÁLEZ TEN

Tutores académicos:
Juan José FONT FERRANDIS
Sergio MACARIO VIVES

Fecha de lectura: 5 de Junio de 2020
Curso académico 2019/2020

Resumen

Este documento recoge el Proyecto de Final de Grado de la asignatura *MT-1030/Prácticas Externas y Proyecto de Final de Grado*, del Grado en Matemática Computacional, cursado en la Universitat Jaume I.

Consta de tres capítulos. El primero introduce la motivación y los objetivos que persigue este proyecto y también desglosa la estructura del trabajo. A continuación, en el segundo capítulo, se presenta el estudio del error en la aproximación de una función continua de una variable mediante redes neuronales activadas usando una función sigmoideal. Por último, en el tercer capítulo se presentan las conclusiones.

Palabras clave

Red neuronal, función sigmoideal acotada, función continua, intervalo compacto.

Abstract

This document details the Final Degree Project of the subject *MT-1030 / External Practices and Final Degree Project*, of the Degree in Computational Mathematics, at the Universitat Jaume I.

It consists of three chapters. The first one introduces the motivation and the goals of this project and also explains the structure of the memory. Next, the second chapter contains the study of the error of the approximation of a continuous map with one variable using neural networks activated by a sigmoideal map. Finally, the third chapter presents the conclusions.

Keywords

Neural network, bounded sigmoideal map, continuous map, compact interval.

Agradecimientos

En primer lugar, me gustaría agradecer a mi familia por el apoyo moral que ha supuesto a lo largo de estos años, ayudándome a levantarme en los peores momentos y por el esfuerzo económico. Además, gracias a ellos pude ir durante nueve meses a Finlandia con el programa Erasmus, lo cual me permitió conocer otro sistema educativo y otros sistemas de evaluación, bastante diferentes a los que tenemos en España.

Han pasado seis años desde que comencé el grado de Matemática Computacional, y a lo largo de este tiempo, he conocido a muchas personas, algunas de las cuales han llegado para quedarse.

Me gustaría agradecer a todos aquellos compañeros que han estado ahí, ayudándome a estudiar, enseñándome a programar o a utilizar el programa LaTeX, formando parte de un trabajo grupal, o simplemente estando conmigo. Sin todos ellos me habría sido muy difícil seguir.

Por otra parte, mis amigos de siempre han sido también un gran apoyo moral, han sabido entenderme cuando estaba agobiada, me han ayudado cuando más lo he necesitado y también hemos compartido juntos muchas horas de biblioteca. Aunque cada uno haya tomado un camino distinto, no ha impedido que sigamos juntos, lo que me hace ver que puedo confiar en ellos siempre.

Por último, quería agradecer a todos los profesores, que durante estos años han hecho una gran labor enseñándonos, ya sea con evaluación continua, trabajos, prácticas, laboratorios o con examen final. En especial, me gustaría agradecer a mis tutores, Juanjo y Sergio, por ayudarme en la realización de este trabajo. Sin su ayuda y su dedicación no habría sido posible.

Índice general

1. Introducción	11
1.1. Contexto y motivación del proyecto	11
2. Memoria TFG	15
2.1. Resultados teóricos	15
2.1.1. Conceptos previos	15
2.1.2. Resultados principales	18
2.2. Ejemplo numérico	37
3. Conclusiones	45
3.1. Ámbito formativo	45
3.2. Ámbito personal	46

Índice de figuras

1.1. Red neuronal	12
2.1. Partición del intervalo $[0,1]$	19
2.2. Complementario de la partición anterior	23
2.3. Partición del intervalo $[0,1]$	26
2.4. Complementario de la partición anterior	31
2.5. Gráfica de la aproximación a $f(x)$ para $n = 15$	41
2.6. Gráfica de la aproximación a $f(x)$ para $n = 25$	41
2.7. Gráfica de la aproximación a $f(x)$ para $n = 50$	42
2.8. Gráfica de la aproximación a $f(x)$ para $n = 100$	42
2.9. Gráfica de la aproximación a $f(x)$ para $n = 200$	43

Capítulo 1

Introducción

1.1. Contexto y motivación del proyecto

Este documento recoge el Proyecto Final de Grado, cuyo objetivo principal es el estudio del error en la aproximación de una función continua de una variable mediante redes neuronales activadas usando una función sigmoïdal.

Hoy en día, vivimos en un constante crecimiento tecnológico y, dentro de estos avances, encontramos el campo de la inteligencia artificial, más en concreto, el campo de las redes de neuronas artificiales o, comúnmente llamadas, redes neuronales. La motivación principal del proyecto es, por tanto, la de conocer y comprender su funcionamiento. Este tema es muy interesante y atractivo, ya que está muy presente en nuestro día a día debido a sus aplicaciones en distintos ámbitos tecnológicos y a su rápido desarrollo.

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. Lo que las redes artificiales pretenden es imitar este comportamiento y llevarlo a un computador. Pero, no solo eso, también pueden formarse a si mismas, es decir, aprender. Esto es un avance, ya que no necesitan ser programadas explícitamente y sobresalen en áreas donde es difícil la programación convencional para extraer soluciones.

Este concepto lleva presente desde la década de los 40, cuando McCulloch y Pitts dieron la primera definición en su artículo [6], 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 general, sus aplicaciones han ido evolucionado hasta la fecha y hoy en día las más importantes son robótica, diagnóstico clínico, análisis de modelos meteorológicos, entre otras muchas.

La siguiente figura 1.1 muestra el esquema de una red neuronal artificial comparado con el de una red neuronal biológica, donde se puede apreciar cómo imita el comportamiento.

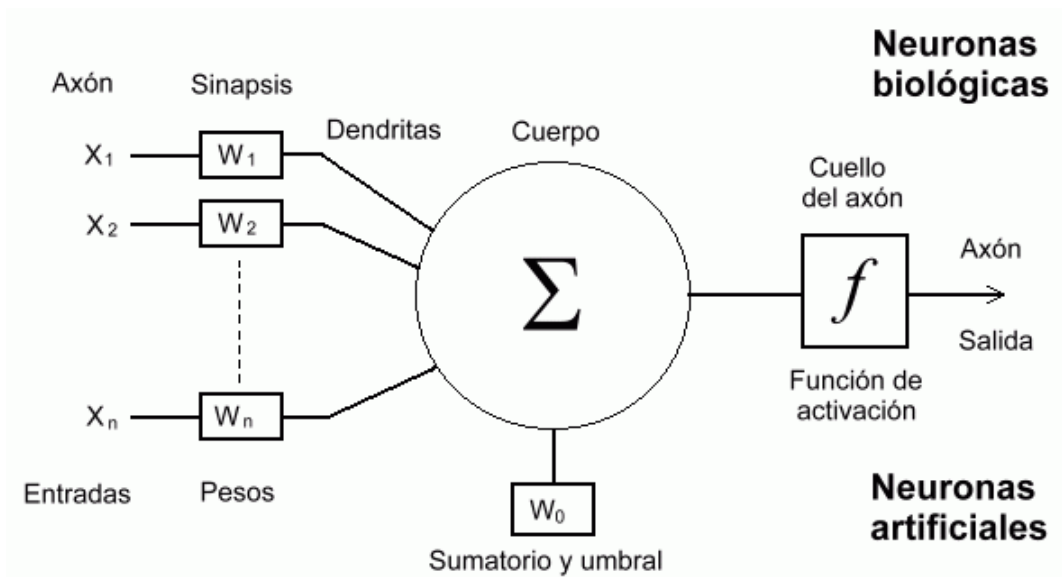


Figura 1.1: Red neuronal

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

$$H(x) = \sum_{i=1}^n a_i \cdot \sigma(x \cdot b_i + c_j)$$

donde los coeficientes $a_j, b_j, c_i \in \mathbb{R}$ son los pesos y la función $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ es la denominada *función de activación*.

Cybenko nos muestra en su estudio, [1], los primeros resultados sobre la aproximación de funciones usando redes neuronales activadas mediante funciones sigmoideas. Otros ejemplos de investigaciones interesantes sobre este tipo de función son las que realizan Hong y Hahm en [3], sobre la aproximación de la red neuronal a una función continua en \mathbb{R} y [4], donde se sugiere

una aproximación constructiva por las redes neuronales con una función sigmoideal utilizando el método de convolución.

Kalman y Kwasny enfatizan en su estudio [5], 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 [7], se sugiere un resultado de densidad por red neuronal con una función sigmoideal usando el teorema de Taylor.

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 3 donde se utilizarán los resultados obtenidos previamente en el Teorema 1 y en el Teorema 2. Para terminar el capítulo presentamos un ejemplo numérico, implementado en FreeMat, que ilustra 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

Memoria TFG

2.1. Resultados teóricos

2.1.1. Conceptos previos

En este documento vamos a centrarnos en las redes neuronales introducidas en el capítulo anterior, es decir, las de tres capas, con las funciones de activación más utilizadas, las funciones sigmoideas. Vamos a definir las y, a continuación, daremos unos ejemplos de las funciones sigmoideas más habituales:

Definición 1 *Una función sigmoidea es una aplicación*

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

que cumple $\lim_{t \rightarrow -\infty} \sigma(t) = 0$ y $\lim_{t \rightarrow +\infty} \sigma(t) = 1$

Las funciones de activación sigmoideas más habituales son las siguientes:

Función de Heaviside o escalón unitario

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

Función lineal a trozos o rampa unidad:

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

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.

Antes de empezar con los resultados, presentamos una serie de definiciones y conceptos que creemos necesarios para entender mejor las demostraciones.

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

Definición 3 Una función $f : \mathbb{R} \rightarrow \mathbb{R}$ 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 $\lim_{n \rightarrow \infty} x_n = x_0$, se tiene que $\lim_{n \rightarrow \infty} f(x_n) = f(x_0)$. Tendremos **continuidad uniforme** cuando δ no dependa de x_0 .

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

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

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

Definición 5 Definimos la norma $\|f\|_{[0,1],\infty}$ como el valor máximo de la función continua f en el intervalo $[0,1]$:

$$\|f\|_{[0,1],\infty} = \max_{x \in [0,1]} (|f(x)|) \quad (2.2)$$

En este documento investigamos la aproximación constructiva de cualquier función continua $f : [0, 1] \rightarrow \mathbb{R}$ mediante las funciones del tipo $\Psi_{n,\sigma}$ y $\Phi_{n,\sigma}$ que, siguiendo la notación de [4] y [5], se definen como:

$$\Psi_{n,\sigma} = \left\{ \sum_{i=1}^n a_0 + a_i \sigma(b_i x + c_i) : a_i, b_i, c_i \in \mathbb{R} \right\}$$

$$\Phi_{n,\sigma} = \left\{ \sum_{i=1}^n a_i \sigma(b_i x + c_i) : a_i, b_i, c_i \in \mathbb{R} \right\}.$$

y, respectivamente, el error de aproximación entre f y Ψ y entre f y Φ como:

$$E_{n,\sigma}^{\Psi}(f) = \inf_{g \in \Psi_{n,\sigma}} \|f - g\|_{[0,1],\infty}$$

$$E_{n,\sigma}^{\Phi}(f) = \inf_{g \in \Phi_{n,\sigma}} \|f - g\|_{[0,1],\infty}.$$

Por último, dada una función continua $f : [0, 1] \rightarrow \mathbb{R}$ definimos su módulo de continuidad como:

$$\omega(f, \delta) = \sup \{ |f(s) - f(t)| : |s - t| \leq \delta; s, t \in [0, 1] \}$$

donde $\delta > 0$. De aquí deducimos que $\omega(f, \delta) \rightarrow 0$ cuando $\delta \rightarrow 0$ por la continuidad uniforme de f en $[0,1]$.

Estas tres últimas definiciones van a ser esenciales para los siguientes resultados.

2.1.2. Resultados principales

Teorema 1 Sean $f : [0, 1] \rightarrow \mathbb{R}$ una función continua y $n \in \mathbb{N}$. Entonces se cumple:

$$E_{n,\sigma}^{\Psi}(f) \leq (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right)$$

donde σ es una función sigmoïdal acotada y $\|\sigma\| = \sup_{x \in \mathbb{R}} |\sigma(x)|$.

Demostración

Fijamos $\epsilon > 0$ y $n \in \mathbb{N}$.

Como f está acotada en $[0,1]$, por ser f continua y $[0,1]$ compacto, existirá $M > 0$ tal que $|f(x)| \leq M \forall x \in [0, 1]$. Por tanto,

$$\omega(f, \alpha) = \sup \{|f(x) - f(y)| : |x - y| \leq \alpha\} \leq 2M, \forall \alpha > 0$$

Como σ es una función sigmoïdal, existe $m_0 \in \mathbb{N}$ de forma que:

$$\begin{aligned} |\sigma(x)| &\leq \frac{\epsilon}{4nM}, & x \leq -m_0 \\ y & \\ |\sigma(x)-1| &\leq \frac{\epsilon}{4nM}, & x \geq m_0 \end{aligned} \tag{2.3}$$

Fijamos $0 < \delta < \frac{1}{2n}$, y definimos la siguiente función $g(x) \in \Psi_{n,\sigma}$:

$$g(x) := f(x_0) + \sum_{i=1}^n (f(x_i) - f(x_{i-1}))\sigma\left(\frac{m_0}{\delta}(x - x_i)\right), \text{ donde } x_i = i/n \text{ con } 0 \leq i \leq n.$$

Nuestro objetivo es probar que $|g(x) - f(x)| \leq (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right), \forall x \in [0, 1]$.

Sea $x \in [x_0, x_0 + \delta) \cup (\bigcup_{j=1}^{n-1} (x_j - \delta, x_j + \delta)) \cup (x_n - \delta, x_n]$. Entonces, existe $k \in \mathbb{N}$ tal que $x \in (x_k - \delta, x_k + \delta)$ donde:

si $k = 0 \rightarrow x \in [x_0, x_0 + \delta) = [0, \delta)$,

si $k = n \rightarrow x \in (x_n - \delta, x_n] = (1 - \delta, 1]$.

Esta partición del intervalo $[0, 1]$ se representa en la siguiente figura, donde la parte coloreada es la que contiene a x :

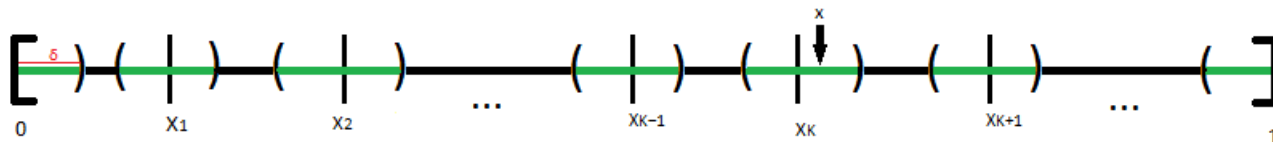


Figura 2.1: Partición del intervalo $[0, 1]$

En esta figura se observa claramente que $(x - x_i) > \delta$ para $1 \leq i \leq k - 1$, $(x - x_i) < -\delta$ para $k + 1 \leq i \leq n$ y que $|x - x_k| < \delta$. Por tanto, si denotamos por $t_i := \frac{m_0}{\delta}(x - x_i)$, $i = 1, \dots, n$, se cumple que:

$$t_i \geq m_0 \quad \text{para} \quad 1 \leq i \leq k - 1$$

$$t_i \leq -m_0 \quad \text{para} \quad k + 1 \leq i \leq n$$

Así pues, deducimos que:

$$|g(x) - f(x)| = \left| f(x_0) + \sum_{i=1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) - f(x) \right| \quad (2.4)$$

Veamos el desarrollo de la expresión:

$$\begin{aligned}
|g(x) - f(x)| &= \left| f(x_0) + (f(x_1) - f(x_0))\sigma(t_1) + (f(x_2) - f(x_1))\sigma(t_2) \right. \\
&\quad + \dots \\
&\quad + (f(x_{k-1}) - f(x_{k-2}))\sigma(t_{k-1}) + (f(x_k) - f(x_{k-1}))\sigma(t_k) + \\
&\quad \left. + \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) - f(x) \right| \\
&= \left| f(x_0) + f(x_1)\sigma(t_1) - f(x_0)\sigma(t_1) + f(x_2)\sigma(t_2) - f(x_1)\sigma(t_2) \right. \\
&\quad + \dots \\
&\quad + f(x_{k-1})\sigma(t_{k-1}) - f(x_{k-2})\sigma(t_{k-1}) + f(x_k)\sigma(t_k) - f(x_{k-1})\sigma(t_k) + \\
&\quad \left. + \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) - f(x) \right| \tag{2.5} \\
&= \left| f(x_1)\sigma(t_1) - f(x_0)\sigma(t_1) - f(x_1) + f(x_0) \right. \\
&\quad + f(x_2)\sigma(t_2) - f(x_1)\sigma(t_2) + f(x_1) - f(x_2) \\
&\quad + \dots \\
&\quad + f(x_{k-1})\sigma(t_{k-1}) - f(x_{k-2})\sigma(t_{k-1}) + f(x_{k-2}) - f(x_{k-1}) \\
&\quad + f(x_k)\sigma(t_k) - f(x_{k-1})\sigma(t_k) + f(x_{k-1}) - f(x_k) \\
&\quad \left. + \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) - f(x) \right|
\end{aligned}$$

Por la desigualdad triangular, podemos concluir que esta expresión es menor o igual que la suma de las siguientes cuatro expresiones:

$$\begin{aligned}
|g(x) - f(x)| &\leq |f(x_1)\sigma(t_1) - f(x_0)\sigma(t_1) - f(x_1) + f(x_0) + f(x_2)\sigma(t_2) - f(x_1)\sigma(t_2) + f(x_1) - f(x_2)| \\
&\quad \mathbf{(1)} + \dots \\
&\quad + |f(x_{k-1})\sigma(t_{k-1}) - f(x_{k-2})\sigma(t_{k-1}) + f(x_{k-2}) - f(x_{k-1})| \\
&\quad \mathbf{(2)} + |f(x_k)\sigma(t_k) - f(x_{k-1})\sigma(t_k) + f(x_{k-1}) - f(x_k)| \\
&\quad \mathbf{(3)} + \left| \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) \right| \\
&\quad \mathbf{(4)} + |f(x_k) - f(x)|
\end{aligned} \tag{2.6}$$

Y vamos a estudiarlas por separado, empezando con (1):

$$\begin{aligned}
&|(f(x_1) - f(x_0))\sigma(t_1) - f(x_1) + f(x_0) + (f(x_2) - f(x_1))\sigma(t_2) - f(x_2) + f(x_1)| \\
&+ \dots \\
&+ |(f(x_{k-1}) - f(x_{k-2}))\sigma(t_{k-1}) - f(x_{k-1}) + f(x_{k-2})| \\
= &|(f(x_1) - f(x_0))(\sigma(t_1) - 1) + (f(x_2) - f(x_1))(\sigma(t_2) - 1)| \\
&+ \dots \\
&+ |(f(x_{k-1}) - f(x_{k-2}))(\sigma(t_{k-1}) - 1)| \\
= &\left| \sum_{i=1}^{k-1} (f(x_i) - f(x_{i-1}))(\sigma(t_i) - 1) \right| \leq T(k-1)\omega\left(f, \frac{1}{n}\right)
\end{aligned} \tag{2.7}$$

donde $T = \frac{\epsilon}{4nM}$ y $\left| f(x_i) - f(x_{i-1}) \right| \leq \omega\left(f, \frac{1}{n}\right)$ ya que $|x_i - x_{i-1}| \leq \frac{i}{n} - \frac{(i-1)}{n} = \frac{1}{n}$ por definición y, además, $\left| \sigma(t_i) - 1 \right| \leq \frac{\epsilon}{4nM}$ por la posición de x en la partición de $[0,1]$.

Veamos (2):

$$\begin{aligned}
|(f(x_k) - f(x_{k-1}))\sigma(t_k) + f(x_{k-1}) - f(x_k)| &\leq |(f(x_k) - f(x_{k-1}))\sigma(t_k)| + |f(x_{k-1}) - f(x_k)| \\
&\leq \|\sigma\| \omega\left(f, \frac{1}{n}\right) + \omega\left(f, \frac{1}{n}\right) = \omega\left(f, \frac{1}{n}\right) (\|\sigma\| + 1)
\end{aligned} \tag{2.8}$$

Esta acotación es parecida a la anterior, pero en este caso la distancia $|x - x_k| < \delta$ ya que hemos tomado x en el intervalo $(x_k - \delta, x_k + \delta)$ y, por tanto, solo vamos a poder asegurar que $|\sigma(t_k)| \leq \sup_{x \in \mathbb{R}} |\sigma(x)| = \|\sigma\|$.

El caso (3) vamos a desarrollarlo y acotarlo de la misma manera que hemos hecho en (1), puesto que $t_i \leq -m_0$ para $i = k + 1, \dots, n$ y, por tanto, $|\sigma(t_i)| < T$:

$$\left| \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) \right| \leq T(n - k) \omega\left(f, \frac{1}{n}\right). \tag{2.9}$$

Por último, ya que $|x - x_k| < \delta$, podemos acotar la expresión (4) por $|f(x_k) - f(x)| \leq \omega(f, \delta)$. Por tanto, si $x \in [x_0, x_0 + \delta) \cup (\bigcup_{j=1}^{n-1} (x_j - \delta, x_j + \delta)) \cup (x_n - \delta, x_n]$, entonces:

$$\begin{aligned}
|g(x) - f(x)| &= \left| f(x_0) + \sum_{i=1}^n (f(x_i) - f(x_{i-1}))\sigma\left(\frac{m_0}{\delta}(x - x_i)\right) - f(x) \right| \\
&\leq T(k - 1) \omega\left(f, \frac{1}{n}\right) + \|\sigma\| \omega\left(f, \frac{1}{n}\right) + \omega\left(f, \frac{1}{n}\right) + T(n - k) \omega\left(f, \frac{1}{n}\right) + \omega(f, \delta) \\
&= T(n - 1) \omega\left(f, \frac{1}{n}\right) + (\|\sigma\| + 1) \omega\left(f, \frac{1}{n}\right) + \omega(f, \delta) \\
&\leq Tn \omega\left(f, \frac{1}{n}\right) + (\|\sigma\| + 1) \omega\left(f, \frac{1}{n}\right) + \omega(f, \delta) \\
&\leq \frac{\epsilon}{4nM} 2Mn + (\|\sigma\| + 1) \omega\left(f, \frac{1}{n}\right) + \omega(f, \delta) \\
&= \frac{\epsilon}{2} + (\|\sigma\| + 1) \omega\left(f, \frac{1}{n}\right) + \omega(f, \delta)
\end{aligned} \tag{2.10}$$

Ahora vamos a estudiar el caso en que $x \in (\bigcup_{j=0}^{n-1} [x_j + \delta, x_{j+1} - \delta])$, donde, como antes, existe $k \in \mathbb{N}$ con $0 \leq k \leq n-1$ tal que $x \in [x_k + \delta, x_{k+1} - \delta]$. En este caso, x está en la zona coloreada siguiente:

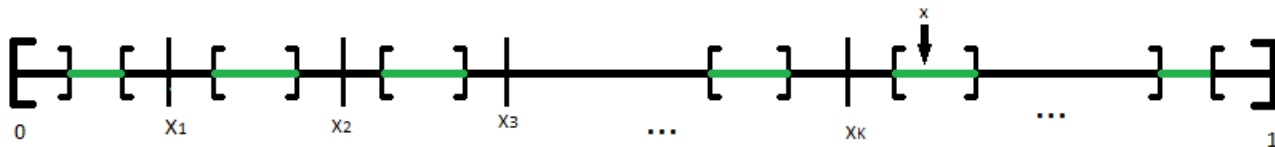


Figura 2.2: Complementario de la partición anterior

La figura 2.2 nos muestra que $x - x_i > \delta$ para $1 \leq i \leq k$ y $x - x_i < -\delta$ para $k+1 \leq i \leq n$. Dicho esto, procedemos a hacer un desarrollo parecido al anterior:

$$\begin{aligned}
|g(x) - f(x)| &= \left| f(x_0) + \sum_{i=1}^n (f(x_i) - f(x_{i-1}))\sigma\left(\frac{m_0}{\delta}(x - x_i)\right) - f(x) \right| \\
&= \left| f(x_0) + (f(x_1) - f(x_0))\sigma(t_1) + (f(x_2) - f(x_1))\sigma(t_2) \right. \\
&\quad + \dots \\
&\quad + (f(x_{k-1}) - f(x_{k-2}))\sigma(t_{k-1}) + (f(x_k) - f(x_{k-1}))\sigma(t_k) \\
&\quad \left. + \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) - f(x) \right| \\
&= \left| f(x_1)\sigma(t_1) - f(x_0)\sigma(t_1) + f(x_0) - f(x_1) + f(x_2)\sigma(t_2) - f(x_1)\sigma(t_2) + f(x_1) - f(x_2) \right. \\
&\quad + \dots \\
&\quad + f(x_{k-1})\sigma(t_{k-1}) - f(x_{k-2})\sigma(t_{k-1}) + f(x_{k-2}) - f(x_{k-1}) + f(x_k)\sigma(t_k) - f(x_{k-1})\sigma(t_k) \\
&\quad + f(x_{k-1}) - f(x_k) \\
&\quad \left. + \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) - f(x) \right|
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
|g(x) - f(x)| &\leq \left| f(x_1)\sigma(t_1) - f(x_0)\sigma(t_1) + f(x_0) - f(x_1) + f(x_2)\sigma(t_2) - f(x_1)\sigma(t_2) + f(x_1) - f(x_2) \right. \\
&\quad \text{(1) + ...} \\
&\quad \left. + f(x_k)\sigma(t_k) - f(x_{k-1})\sigma(t_k) + f(x_{k-1}) - f(x_k) \right| \\
&\quad \text{(2) + } \left| \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) \right| \\
&\quad \text{(3) + } \left| f(x_k) - f(x) \right|
\end{aligned} \tag{2.12}$$

Estos casos ya han sido estudiados, por tanto tenemos que si $x \in (\bigcup_{j=0}^{n-1} [x_j + \delta, x_{j+1} - \delta])$, entonces:

$$\begin{aligned}
|g(x) - f(x)| &= \left| \sum_{i=1}^k (f(x_i) - f(x_{i-1}))(\sigma(t_i) - 1) \right| + \left| \sum_{i=k+1}^n (f(x_i) - f(x_{i-1}))\sigma(t_i) \right| + \left| f(x_k) - f(x) \right| \\
&\leq Tk\omega\left(f, \frac{1}{n}\right) + T(n-k)\omega\left(f, \frac{1}{n}\right) + \omega\left(f, \frac{1}{n}\right) \\
&\leq Tn\omega\left(f, \frac{1}{n}\right) + \omega\left(f, \frac{1}{n}\right) \\
&\leq \frac{\epsilon}{2} + \omega\left(f, \frac{1}{n}\right) \leq \frac{\epsilon}{2} + (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right) + \omega(f, \delta)
\end{aligned} \tag{2.13}$$

Recapitulando desde el principio, tenemos que, para todo $x \in [0, 1]$

$$|g(x) - f(x)| \leq \frac{\epsilon}{2} + (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right) + \omega(f, \delta) \tag{2.14}$$

Por tanto, $\|g - f\|_\infty \leq \frac{\epsilon}{2} + (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right) + \omega(f, \delta)$ y, en consecuencia,

$$E_{n,\sigma}^\Psi(f) = \inf_{h \in \Psi_{n,\sigma}} \|f - h\|_\infty \leq \|g - f\|_\infty \leq \frac{\epsilon}{2} + (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right) + \omega(f, \delta) \quad (2.15)$$

Puesto que $\epsilon > 0$ y $\delta > 0$ son arbitrarios, podemos concluir que:

$$E_{n,\sigma}^\Psi(f) \leq (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right) \quad (2.16)$$

□

Teorema 2 Sean h una función constante (distinta de cero) en el intervalo $[0, 1]$ y $n \in \mathbb{N}$. Entonces se cumple:

$$E_{2n,\sigma}^\Phi(h) \leq \frac{c}{n}$$

donde σ es una función sigmoideal acotada y c es una constante dependiente de h y de σ .

Demostración

Fijamos $\epsilon > 0$ y $n \in \mathbb{N}$.

Como en el teorema anterior, existe $m_0 \in \mathbb{N}$ que cumple:

$$\begin{aligned} |\sigma(x)| &\leq \frac{\epsilon}{2n}, & x \leq -m_0 \\ &y \\ |\sigma(x) - 1| &\leq \frac{\epsilon}{2n}, & x \geq m_0 \end{aligned} \quad (2.17)$$

Sea $h(x) = d \neq 0$, $\forall x \in [0, 1]$; $x_i = \frac{i}{n}$, $0 \leq i \leq n$. Definimos la siguiente función para cualquier $0 < \delta < \frac{1}{2n}$:

$$p(x) = \frac{d}{n} \sum_{j=1}^{2n} \sigma\left(c_j \frac{m_0}{\delta} (x - x_j)\right)$$

donde $c_j = 1$ para $1 \leq j \leq n$ y $c_j = -1$, $x_j = x_{j-n}$ para $n+1 \leq j \leq 2n$. Es decir, $p(x) \in \Phi_{2n,\sigma}$.

Con las características descritas, podemos reescribir la anterior función como sigue:

$$p(x) = \frac{d}{n} \sum_{j=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_j) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_j) \right) \right)$$

Si $x \in [x_0, x_0 + \delta) \cup (\bigcup_{j=1}^{n-1} (x_j - \delta, x_j + \delta)) \cup (x_n - \delta, x_n]$ entonces, existe $k \in \mathbb{N}$, con $0 \leq k \leq n$, tal que $x \in (x_n - \delta, x_k + \delta)$ donde:

si $k = 0 \rightarrow x \in [x_0, x_0 + \delta) = [0, \delta)$

si $k = n \rightarrow x \in (x_n - \delta, x_n] = (1 - \delta, 1]$

Es decir, tenemos la misma partición anterior para $[0, 1]$ y, por tanto, se sigue cumpliendo $x - x_i > \delta$ para $1 \leq i \leq k - 1$ y $x - x_i < -\delta$ para $k + 1 \leq i \leq n$:

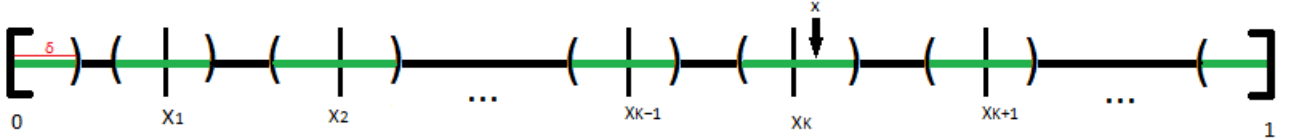


Figura 2.3: Partición del intervalo $[0, 1]$

Por tanto tenemos:

$$|p(x) - h(x)| = \left| \frac{d}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right) - d \right|$$

De aquí en adelante volvemos a denotar $t_i = \frac{m_0}{\delta} (x - x_i)$, $i = 1, \dots, n$ como se ha hecho en el teorema anterior, por lo cual:

$$t_i \geq m_0, \quad \text{para } 1 \leq i \leq k - 1$$

$$t_i \leq -m_0, \quad \text{para } k + 1 \leq i \leq n$$

Desarrollamos el sumatorio, añadimos $\left(\frac{d}{n} - \frac{d}{n}\right)$ por cada iteración y reagrupamos para extraer

$\frac{d}{n}$ como factor común:

$$\begin{aligned}
|p(x) - h(x)| &= \left| \frac{d}{n} [\sigma(t_1) + \sigma(-t_1)] + \frac{d}{n} [\sigma(t_2) + \sigma(-t_2)] + \frac{d}{n} [\sigma(t_3) + \sigma(-t_3)] \right. \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{k-1}) + \sigma(-t_{k-1})] + \frac{d}{n} [\sigma(t_k) + \sigma(-t_k)] + \frac{d}{n} [\sigma(t_{k+1}) + \sigma(-t_{k+1})] \\
&\quad + \dots \\
&\quad \left. + \frac{d}{n} [\sigma(t_{n-1}) + \sigma(-t_{n-1})] + \frac{d}{n} [\sigma(t_n) + \sigma(-t_n)] - d \right| \\
&= \left| \frac{d}{n} [\sigma(t_1) + \sigma(-t_1)] + \left(\frac{d}{n} - \frac{d}{n}\right) + \frac{d}{n} [\sigma(t_2) + \sigma(-t_2)] + \left(\frac{d}{n} - \frac{d}{n}\right) \right. \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{k-1}) + \sigma(-t_{k-1})] + \left(\frac{d}{n} - \frac{d}{n}\right) + \frac{d}{n} [\sigma(t_k) + \sigma(-t_k)] + \left(\frac{d}{n} - \frac{d}{n}\right) \\
&\quad + \dots \\
&\quad \left. + \frac{d}{n} [\sigma(t_{n-1}) + \sigma(-t_{n-1})] + \left(\frac{d}{n} - \frac{d}{n}\right) + \frac{d}{n} [\sigma(t_n) + \sigma(-t_n)] + \left(\frac{d}{n} - \frac{d}{n}\right) - d \right| \\
&= \left| \frac{d}{n} [\sigma(t_1) + \sigma(-t_1) - 1] + \frac{d}{n} + \frac{d}{n} [\sigma(t_2) + \sigma(-t_2) - 1] + \frac{d}{n} \right. \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{k-1}) + \sigma(-t_{k-1}) - 1] + \frac{d}{n} + \frac{d}{n} [\sigma(t_k) + \sigma(-t_k) - 1] + \frac{d}{n} \\
&\quad + \dots \\
&\quad \left. + \frac{d}{n} [\sigma(t_{n-1}) + \sigma(-t_{n-1}) - 1] + \frac{d}{n} + \frac{d}{n} [\sigma(t_n) + \sigma(-t_n) - 1] + \frac{d}{n} - d \right|
\end{aligned} \tag{2.18}$$

$$\begin{aligned}
|p(x) - h(x)| = & \left| \frac{d}{n} \left(\left[\sigma(t_1) + \sigma(-t_1) - 1 \right] + \left[\sigma(t_2) + \sigma(-t_2) - 1 \right] + \left[\sigma(t_3) + \sigma(-t_3) - 1 \right] \right. \right. \\
& + \dots \\
& + \left. \left[\sigma(t_{k-1}) + \sigma(-t_{k-1}) - 1 \right] + \left[\sigma(t_k) + \sigma(-t_k) - 1 \right] + \left[\sigma(t_{k+1}) + \sigma(-t_{k+1}) - 1 \right] \right. \\
& + \dots \\
& \left. + \left[\sigma(t_{n-1}) + \sigma(-t_{n-1}) - 1 \right] + \left[\sigma(t_n) + \sigma(-t_n) - 1 \right] \right) + n \frac{d}{n} - d \Big|
\end{aligned} \tag{2.19}$$

Por la desigualdad triangular, esta expresión es menor o igual que la suma de las siguientes tres expresiones:

$$\begin{aligned}
|p(x) - h(x)| \leq & \left| \frac{d}{n} \left(\left[\sigma(t_1) + \sigma(-t_1) - 1 \right] + \left[\sigma(t_2) + \sigma(-t_2) - 1 \right] + \left[\sigma(t_3) + \sigma(-t_3) - 1 \right] \right. \right. \\
& \mathbf{(1)} + \dots \\
& \left. \left. + \left[\sigma(t_{k-1}) + \sigma(-t_{k-1}) - 1 \right] \right) \right| \\
& \mathbf{(2)} + \left| \frac{d}{n} \left[\sigma(t_k) + \sigma(-t_k) - 1 \right] \right| \\
& \mathbf{(3)} + \left| \frac{d}{n} \left(\left[\sigma(t_{k+1}) + \sigma(-t_{k+1}) - 1 \right] + \dots + \left[\sigma(t_{n-1}) + \sigma(-t_{n-1}) - 1 \right] \right. \right. \\
& \left. \left. + \left[\sigma(t_n) + \sigma(-t_n) - 1 \right] \right) \right|
\end{aligned} \tag{2.20}$$

Veamos (3), donde tendremos en cuenta que $t_i \leq -m_o$, para $k+1 \leq i \leq n$:

$$\begin{aligned}
& \left| \frac{d}{n} \left(\left[\sigma(t_{k+1}) + (\sigma(-t_{k+1}) - 1) \right] + \dots + \left[\sigma(t_{n-1}) + (\sigma(-t_{n-1}) - 1) \right] \right. \right. \\
& \left. \left. + \left[\sigma(t_n) + (\sigma(-t_n) - 1) \right] \right) \right| \\
&= \left| \frac{d}{n} \sum_{j=k+1}^n \left(\sigma(t_j) + (\sigma(-t_j) - 1) \right) \right| \\
&= \frac{|d|}{n} \left| \sum_{j=k+1}^n \left(\sigma(t_j) + (\sigma(-t_j) - 1) \right) \right| \\
&\leq \frac{|d|}{n} (n-k) \left(\frac{\epsilon}{2n} + \frac{\epsilon}{2n} \right) \\
&= \frac{|d|}{n} \frac{2\epsilon(n-k)}{2n} = \frac{|d|}{n} \frac{(n-k)\epsilon}{n}
\end{aligned} \tag{2.23}$$

En este caso las acotaciones son las mismas que en (1) pero teniendo en cuenta el cambio de signo. Recapitulando, tenemos:

$$\begin{aligned}
|p(x) - h(x)| &= \left| \frac{d}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right) - d \right| \\
&\leq \frac{|d|}{n} \left(\frac{(n-k)\epsilon}{n} + (2\|\sigma\| + 1) + \frac{(k-1)\epsilon}{n} \right) \\
&= \frac{|d|}{n} \left(\frac{(n-k+k-1)\epsilon}{n} + (2\|\sigma\| + 1) \right) = \frac{|d|}{n} \left(\frac{(n-1)\epsilon}{n} + (2\|\sigma\| + 1) \right) \\
&\leq \frac{|d|}{n} \left(\frac{n\epsilon}{n} + (2\|\sigma\| + 1) \right) = \frac{|d|\epsilon}{n} + \frac{|d|(2\|\sigma\| + 1)}{n}
\end{aligned} \tag{2.24}$$

Ahora vamos a tomar $x \in (\bigcup_{j=0}^{n-1} [x_j + \delta, x_{j+1} - \delta])$, donde, como antes, existirá $k \in \mathbb{N}$ con $0 \leq k \leq n - 1$ tal que $x \in [x_k + \delta, x_{k+1} - \delta]$. En este caso la representación de puntos que nos queda es la siguiente:

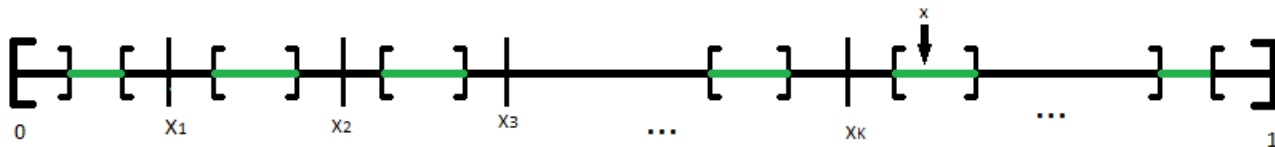


Figura 2.4: Complementario de la partición anterior

Por tanto, tenemos:

$$|p(x) - h(x)| = \left| \frac{d}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right) - d \right|$$

donde volvemos a denotar $t_i = \frac{m_0}{\delta} (x - x_i)$, $i = 1, \dots, n$, por lo que:

$$t_i \geq m_0, \quad \text{para } 1 \leq i \leq k$$

$$t_i \leq -m_0, \quad \text{para } k + 1 \leq i \leq n$$

Desarrollamos el sumatorio como hemos hecho antes:

$$\begin{aligned} |p(x) - h(x)| &= \left| \frac{d}{n} \left[\sigma(t_1) + \sigma(-t_1) \right] + \frac{d}{n} \left[\sigma(t_2) + \sigma(-t_2) \right] + \frac{d}{n} \left[\sigma(t_3) + \sigma(-t_3) \right] \right. \\ &\quad + \dots \\ &\quad + \frac{d}{n} \left[\sigma(t_{k-1}) + \sigma(-t_{k-1}) \right] + \frac{d}{n} \left[\sigma(t_k) + \sigma(-t_k) \right] + \frac{d}{n} \left[\sigma(t_{k+1}) + \sigma(-t_{k+1}) \right] \\ &\quad + \dots \\ &\quad \left. + \frac{d}{n} \left[\sigma(t_{n-1}) + \sigma(-t_{n-1}) \right] + \frac{d}{n} \left[\sigma(t_n) + \sigma(-t_n) \right] - d \right| \end{aligned} \tag{2.25}$$

Añadimos $\left(\frac{d}{n} - \frac{d}{n}\right)$ por cada iteración:

$$\begin{aligned}
|p(x) - h(x)| &= \left| \frac{d}{n} [\sigma(t_1) + \sigma(-t_1)] + \left(\frac{d}{n} - \frac{d}{n}\right) + \frac{d}{n} [\sigma(t_2) + \sigma(-t_2)] + \left(\frac{d}{n} - \frac{d}{n}\right) \right. \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{k-1}) + \sigma(-t_{k-1})] + \left(\frac{d}{n} - \frac{d}{n}\right) + \frac{d}{n} [\sigma(t_k) + \sigma(-t_k)] + \left(\frac{d}{n} - \frac{d}{n}\right) \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{n-1}) + \sigma(-t_{n-1})] + \left(\frac{d}{n} - \frac{d}{n}\right) + \frac{d}{n} [\sigma(t_n) + \sigma(-t_n)] + \left(\frac{d}{n} - \frac{d}{n}\right) - d \left| \right. \\
&= \left| \frac{d}{n} [\sigma(t_1) + \sigma(-t_1) - 1] + \frac{d}{n} + \frac{d}{n} [\sigma(t_2) + \sigma(-t_2) - 1] + \frac{d}{n} \right. \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{k-1}) + \sigma(-t_{k-1}) - 1] + \frac{d}{n} + \frac{d}{n} [\sigma(t_k) + \sigma(-t_k) - 1] + \frac{d}{n} \\
&\quad + \dots \\
&\quad + \frac{d}{n} [\sigma(t_{n-1}) + \sigma(-t_{n-1}) - 1] + \frac{d}{n} + \frac{d}{n} [\sigma(t_n) + \sigma(-t_n) - 1] + \frac{d}{n} - d \left| \right. \\
&= \left| \frac{d}{n} \left([\sigma(t_1) + \sigma(-t_1) - 1] + [\sigma(t_2) + \sigma(-t_2) - 1] + [\sigma(t_3) + \sigma(-t_3) - 1] \right. \right. \\
&\quad + \dots \\
&\quad + [\sigma(t_{k-1}) + \sigma(-t_{k-1}) - 1] + [\sigma(t_k) + \sigma(-t_k) - 1] + [\sigma(t_{k+1}) + \sigma(-t_{k+1}) - 1] \\
&\quad + \dots \\
&\quad \left. \left. + [\sigma(t_{n-1}) + \sigma(-t_{n-1}) - 1] + [\sigma(t_n) + \sigma(-t_n) - 1] \right) + n \frac{d}{n} - d \right|
\end{aligned} \tag{2.26}$$

Por la desigualdad triangular, la expresión anterior es menor o igual a la suma de las siguientes dos expresiones:

$$\begin{aligned}
|p(x) - h(x)| &\leq \left| \frac{d}{n} \left(\left[\sigma(t_1) + \sigma(-t_1) - 1 \right] + \left[\sigma(t_2) + \sigma(-t_2) - 1 \right] + \left[\sigma(t_3) + \sigma(-t_3) - 1 \right] \right. \right. \\
&\quad \text{(1) + ...} \\
&\quad \left. \left. + \left[\sigma(t_k) + \sigma(-t_k) - 1 \right] \right) \right| \\
&\quad \text{(2) + } \left| \frac{d}{n} \left(\left[\sigma(t_{k+1}) + \sigma(-t_{k+1}) - 1 \right] + \dots + \left[\sigma(t_{n-1}) + \sigma(-t_{n-1}) - 1 \right] \right. \right. \\
&\quad \left. \left. + \left[\sigma(t_n) + \sigma(-t_n) - 1 \right] \right) \right|
\end{aligned} \tag{2.27}$$

Primero vemos (1):

$$\begin{aligned}
&\left| \frac{d}{n} \left(\left[\sigma(t_1) + \sigma(-t_1) - 1 \right] + \left[\sigma(t_2) + \sigma(-t_2) - 1 \right] + \left[\sigma(t_3) + \sigma(-t_3) - 1 \right] \right. \right. \\
&\quad + \dots \\
&\quad \left. \left. + \left[\sigma(t_k) + \sigma(-t_k) - 1 \right] \right) \right| \\
&= \left| \frac{d}{n} \sum_{j=1}^k \left(\sigma(t_j) + \sigma(-t_j) - 1 \right) \right| = \frac{|d|}{n} \left| \sum_{j=1}^k \left((\sigma(t_j) - 1) + \sigma(-t_j) \right) \right| \\
&\leq \frac{|d|k}{n} \left(\frac{\epsilon}{2n} + \frac{\epsilon}{2n} \right) = \frac{|d|k}{n} \frac{2\epsilon}{2n} = \frac{|d|}{n} \frac{k\epsilon}{n}
\end{aligned} \tag{2.28}$$

porque, por una parte, $\left| \sigma(t_i) - 1 \right| \leq \frac{\epsilon}{2n}$ ya que $t_i \geq m_0$. Y si cambiamos el signo, tenemos que $-t_i \leq -m_0$, por lo que, también se cumple $\left| \sigma(-t_i) \right| \leq \frac{\epsilon}{2n}$.

Veamos (2) y tendremos en cuenta que $t_i \leq -m_o$, para $k+1 \leq i \leq n$:

$$\begin{aligned}
& \left| \frac{d}{n} \left(\left[\sigma(t_{k+1}) + (\sigma(-t_{k+1}) - 1) \right] + \dots + \left[\sigma(t_{n-1}) + (\sigma(-t_{n-1}) - 1) \right] \right. \right. \\
& \left. \left. + \left[\sigma(t_n) + (\sigma(-t_n) - 1) \right] \right) \right| \\
& = \left| \frac{d}{n} \sum_{j=k+1}^n \left(\sigma(t_j) + (\sigma(-t_j) - 1) \right) \right| = \frac{|d|}{n} \left| \sum_{j=k+1}^n \left(\sigma(t_j) + (\sigma(-t_j) - 1) \right) \right| \\
& \leq \frac{|d|}{n} (n-k) \left(\frac{\epsilon}{2n} + \frac{\epsilon}{2n} \right) = \frac{|d|}{n} \frac{2\epsilon(n-k)}{2n} = \frac{|d|}{n} \frac{(n-k)\epsilon}{n}
\end{aligned} \tag{2.29}$$

Es decir, tenemos un resultado parecido al obtenido en (1). En este caso las acotaciones son las mismas pero teniendo en cuenta el cambio de signo. Recapitulando, tenemos:

$$\begin{aligned}
|p(x) - h(x)| & = \left| \frac{d}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right) - d \right| \\
& \leq \frac{|d|}{n} \left(\frac{(n-k)\epsilon}{n} + \frac{k\epsilon}{n} \right) = \frac{|d|}{n} \left(\frac{(n-k+k)\epsilon}{n} \right) = \frac{|d|}{n} \left(\frac{n\epsilon}{n} \right) \\
& = \frac{|d|\epsilon}{n} \leq \frac{|d|\epsilon}{n} + \frac{|d|(2\|\sigma\| + 1)}{n}
\end{aligned} \tag{2.30}$$

Es decir, hemos demostrado que $|p(x) - h(x)| \leq \frac{|d|\epsilon}{n} + \frac{|d|(2\|\sigma\| + 1)}{n}$ para todo $x \in [0, 1]$. Por tanto queda demostrado que:

$$E_{2n,\sigma}^\Phi(h) = \inf_{g \in \Phi_{2n,\sigma}} \|h - g\|_\infty \leq \|p - h\|_\infty \leq \frac{|d|\epsilon}{n} + \frac{|d|(2\|\sigma\| + 1)}{n}. \tag{2.31}$$

Como $\epsilon > 0$ es un valor arbitrario, entonces:

$$E_{2n,\sigma}^\Phi(h) \leq \frac{c}{n} \tag{2.32}$$

con $c = |d|(2\|\sigma\| + 1)$.

□

Teorema 3 Sean $f : [0, 1] \rightarrow \mathbb{R}$ una función continua en $[0, 1]$ y $n \in \mathbb{N}$. Entonces:

$$E_{3n, \sigma}^{\Phi}(f) \leq (\|\sigma\| + 1)\omega\left(f, \frac{1}{n}\right) + \frac{c}{n}$$

donde σ es una función sigmoïdal acotada con $\|\sigma\| = \sup_{x \in \mathbb{R}} |\sigma(x)|$ y c es una constante dependiente de f y de σ .

Demostración

Fijamos $\epsilon > 0$ y $n \in \mathbb{N}$.

Como f es continua, entonces está acotada en el compacto $[0, 1]$, es decir, existirá $M > 0$ tal que $|f(x)| \leq M \forall x \in [0, 1]$. Por tanto,

$$\omega(f, \alpha) = \sup \{|f(x) - f(y)| : |x - y| \leq \alpha\} \leq 2M, \forall \alpha > 0$$

Como en los teoremas anteriores, existe $m_0 \in \mathbb{N}$ que cumple:

$$\begin{aligned} |\sigma(x)| &\leq \frac{\epsilon}{4nM}, & x \leq -m_0 \\ y & & \\ |\sigma(x) - 1| &\leq \frac{\epsilon}{4nM}, & x \geq m_0 \end{aligned} \tag{2.33}$$

Sea $x_i = \frac{i}{n}$ para $0 \leq i \leq n$. Fijamos δ con $0 < \delta < \frac{1}{2n}$ y definimos:

$$q(x) = \sum_{i=1}^{3n} a_i \sigma\left(b_i \frac{m_0}{\delta} (x - x_i)\right) \tag{2.34}$$

donde

$$a_i = \begin{cases} f(x_i) - f(x_{i-1}), & 1 \leq i \leq n \\ \frac{f(x_0)}{n}, & n + 1 \leq i \leq 3n \end{cases}$$

$$b_i = \begin{cases} 1, & 1 \leq i \leq 2n \\ -1, & 2n + 1 \leq i \leq 3n \end{cases}$$

$$x_i = \begin{cases} x_{i-n}, & n + 1 \leq i \leq 2n \\ x_{i-2n}, & 2n + 1 \leq i \leq 3n \end{cases}$$

Entonces, $q(x) \in \Phi_{3n,\sigma}$ y, con las características descritas, la podemos reescribir de la siguiente manera:

$$q(x) = \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \frac{f(x_0)}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right)$$

Definimos $q_1 \in \Psi_{n,\sigma}$ como sigue:

$$q_1(x) = f(x_0) + \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \sigma \left(\frac{m_0}{\delta} (x - x_i) \right)$$

Por tanto, $\forall x \in [0, 1]$, tenemos:

$$|q(x) - f(x)| = |q(x) - q_1(x) + q_1(x) - f(x)| \leq \underbrace{|q(x) - q_1(x)|}_{(1)} + \underbrace{|q_1(x) - f(x)|}_{(2)} \quad (2.35)$$

Vamos a ver (1):

$$\begin{aligned} |q(x) - q_1(x)| &= \left| \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \sigma \left(\frac{m_0}{\delta} (x - x_i) \right) \right. \\ &\quad \left. + \frac{f(x_0)}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right) \right. \\ &\quad \left. - f(x_0) - \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \sigma \left(\frac{m_0}{\delta} (x - x_i) \right) \right| \quad (2.36) \\ &= \left| \frac{f(x_0)}{n} \sum_{i=1}^n \left(\sigma \left(\frac{m_0}{\delta} (x - x_i) \right) + \sigma \left(-\frac{m_0}{\delta} (x - x_i) \right) \right) - f(x_0) \right| \end{aligned}$$

Este resultado lo hemos estudiado en el Teorema 2, por tanto, aplicando este teorema:

$$\|q - q_1\|_{[0,1],\infty} \leq \frac{|f(x_0)|\epsilon}{n} + \frac{c}{n}, \quad \text{donde } c = |f(x_0)|(2\|\sigma\| + 1) \quad (2.37)$$

Veamos (2):

$$|q_1(x) - f(x)| = \left| f(x_0) + \sum_{i=1}^n (f(x_i) - f(x_{i-1})) \sigma \left(\frac{m_0}{\delta} (x - x_i) \right) - f(x) \right| \quad (2.38)$$

Este resultado lo hemos estudiado en el Teorema 1, por tanto, aplicando este teorema:

$$\|q_1 - f\|_{[0,1],\infty} \leq \frac{\epsilon}{2} + (\|\sigma\| + 1) \omega \left(f, \frac{1}{n} \right) + \omega(f, \delta) \quad (2.39)$$

Concluyendo, tenemos que:

$$E_{3n,\sigma}^{\Phi}(f) = \inf_{h \in \Phi_{3n,\sigma}} \|f - h\|_{\infty} \leq \|q - f\|_{\infty} \leq \frac{\epsilon}{2} + (\|\sigma\| + 1) \omega \left(f, \frac{1}{n} \right) + \omega(f, \delta) + \frac{|f(x_0)|\epsilon}{n} + \frac{c}{n}$$

y, teniendo en cuenta que $\delta > 0$ y $\epsilon > 0$ son arbitrarios, podemos concluir que:

$$E_{3n,\sigma}^{\Phi}(f) \leq (\|\sigma\| + 1) \omega \left(f, \frac{1}{n} \right) + \frac{c}{n}$$

donde $c = |f(x_0)|(2\|\sigma\| + 1)$ y $\|\sigma\| = \sup_{x \in \mathbb{R}} |\sigma(x)|$.

□

2.2. Ejemplo numérico

Para ilustrar los resultados obtenidos anteriormente, vamos a aproximar una función polinómica, implementando en FreeMat las técnicas utilizadas en la sección precedente. Para ello se considera:

$$f(x) = -8x^3 + 18x^2 - 9x + 1$$

como función a aproximar y la función logística como función de activación:

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

También elegimos $\delta = 0.0001 > 0$ y $m_0 = 100$, ya que $|\sigma(x)| \leq 3,720076 \times 10^{-44}$ para $x \leq 100$ y $|\sigma(x) - 1| \leq 3,720076 \times 10^{-44}$ para $x \geq 100$. La red neuronal que vamos a utilizar es la que aparece en la prueba del Teorema 3, es decir:

$$q(x) = \sum_{i=1}^{3n} a_i \sigma \left(b_i \frac{m_0}{\delta} (x - x_i) \right) \quad (2.40)$$

donde

$$a_i = \begin{cases} f(x_i) - f(x_{i-1}), & 1 \leq i \leq n \\ \frac{f(x_0)}{n}, & n + 1 \leq i \leq 3n \end{cases}$$

$$b_i = \begin{cases} 1, & 1 \leq i \leq 2n \\ -1, & 2n + 1 \leq i \leq 3n \end{cases} \quad (2.41)$$

$$x_i = \begin{cases} x_{i-n}, & n + 1 \leq i \leq 2n \\ x_{i-2n}, & 2n + 1 \leq i \leq 3n \end{cases}$$

Los tres siguientes bloques de código muestran la definición de los coeficientes a_i , b_i y los puntos x_i , que aparecen en la definición de la función de activación:

```
% Devuelve el valor de ai segun la posicion del indice
function a_i = ai(indice, n)
    if indice <= n
        a_i = f(indice/n) - f((indice-1)/n);
    else
        a_i = 1/n;
    end
end
```

```
% Devuelve el valor de bi segun la posicion del indice
function b_i = bi(indice, n)
    if indice <= 2*n
        b_i = 1;
    else
        b_i = -1;
    end
end
```

```

% Devuelve el valor de xi segun la posicion del indice
function x_i = xi(indice , n)
    if i <= n
        x_i = indice/n;
    else if i >= n+1 and i <= 2n
        x_i = (indice-n)/n;
    else
        x_i = (indice - 2*n)/n;
    end
end
end

```

Estas dos funciones se utilizan para calcular el valor de un punto en el polinomio $f(x)$ y en la red neuronal $q(x)$:

```

% Devuelve el valor del polinomio f(x) para un valor concreto
function y = f(x)
    y = -8*x.^3+18*x.^2-9*x+1;
end

```

```

% Devuelve el valor de q(x) para un valor concreto
function res = q(x, n, mo, delta)
    res = 0;
    for i = 1:3*n
        res = res + ai(i,n)*(1/(1+exp(-bi(i,n)*(mo/delta)*
            (x-xi(i,n))))));
    end
end
end

```

El siguiente código devuelve una aproximación al valor de la norma $\|q - f\|_{[0,1],\infty}$ 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
end

```

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

```

%VARIABLES FIJAS
mo = 100;
delta = 0.0001;
n = ;
x = linspace(0,1);

%CONSTRUIR POLINOMIO f(x)
a3 = -8;
a2 = 18;
a1 = -9;
a0 = 1;

f_x = a3*x.^3+a2*x.^2+a1*x+a0;

%CONSTRUIR q(x)
q_x = 0;
for i = 1:3*n
    q_x = q_x + ai(i,n)*(1/(1+exp(-bi(i,n)*(mo/delta)*(x-xi(i,n))))));
end

%CALCULAR ERROR MAXIMO EN LA APROXIMACION
k = 0;
erroraprox = [];

    for y = linspace(0,1,200)
        k = k+1;
        erroraprox(k) = abs(f(y)-q(y,n, mo, delta));
    end

%MOSTRAR RESULTADOS
disp(max_error(erroraprox,200));
plot(x,f_x,'c',x,q_x,'k');

```

A continuación, las siguientes gráficas, obtenidas con el código anterior, ilustran como mejora la aproximación al incrementar el valor de n, concretamente para n= 15, n= 25, n= 50, n= 100 y n= 200.

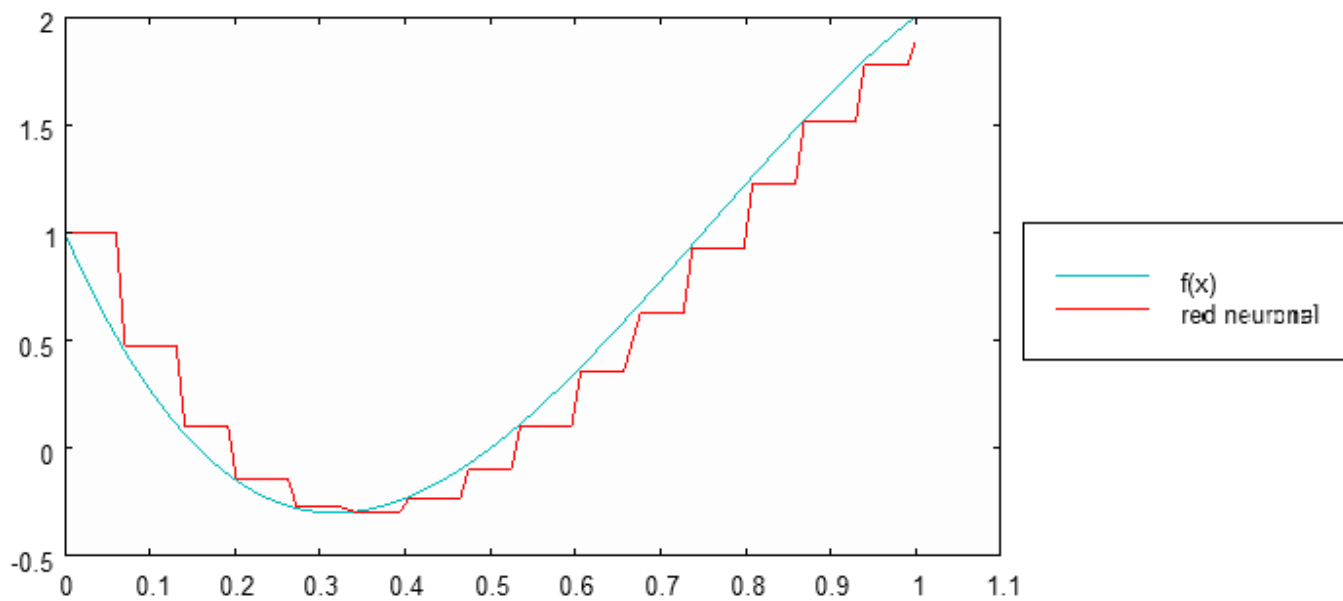


Figura 2.5: Gráfica de la aproximación a $f(x)$ para $n = 15$

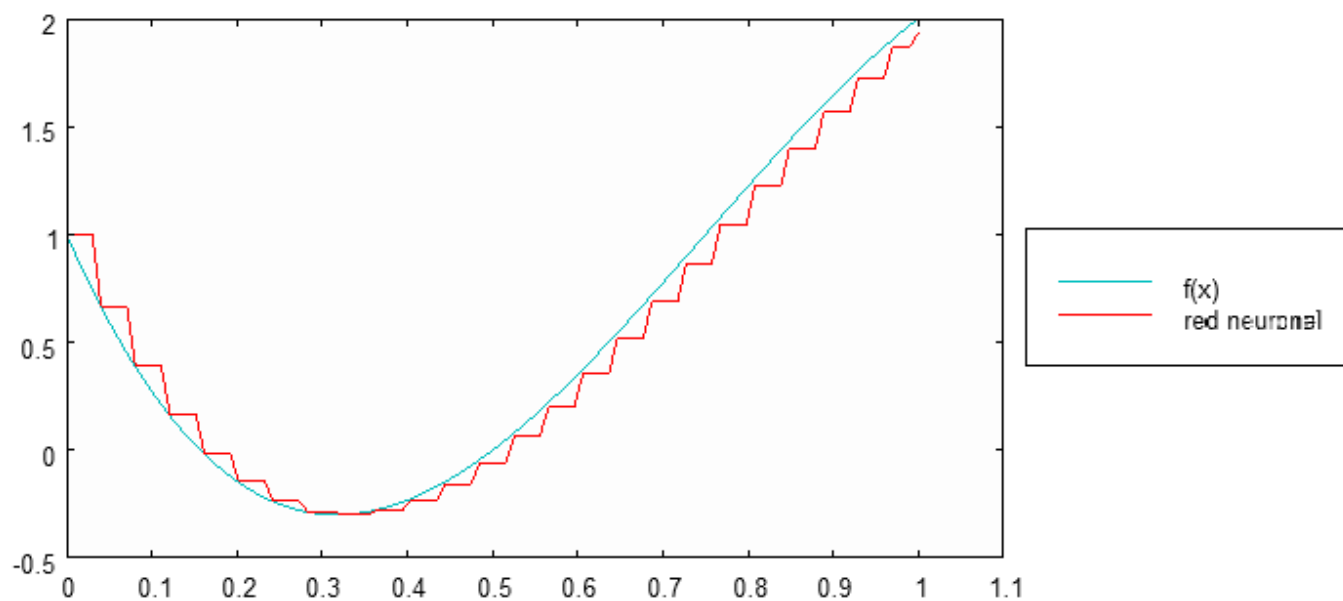


Figura 2.6: Gráfica de la aproximación a $f(x)$ para $n = 25$

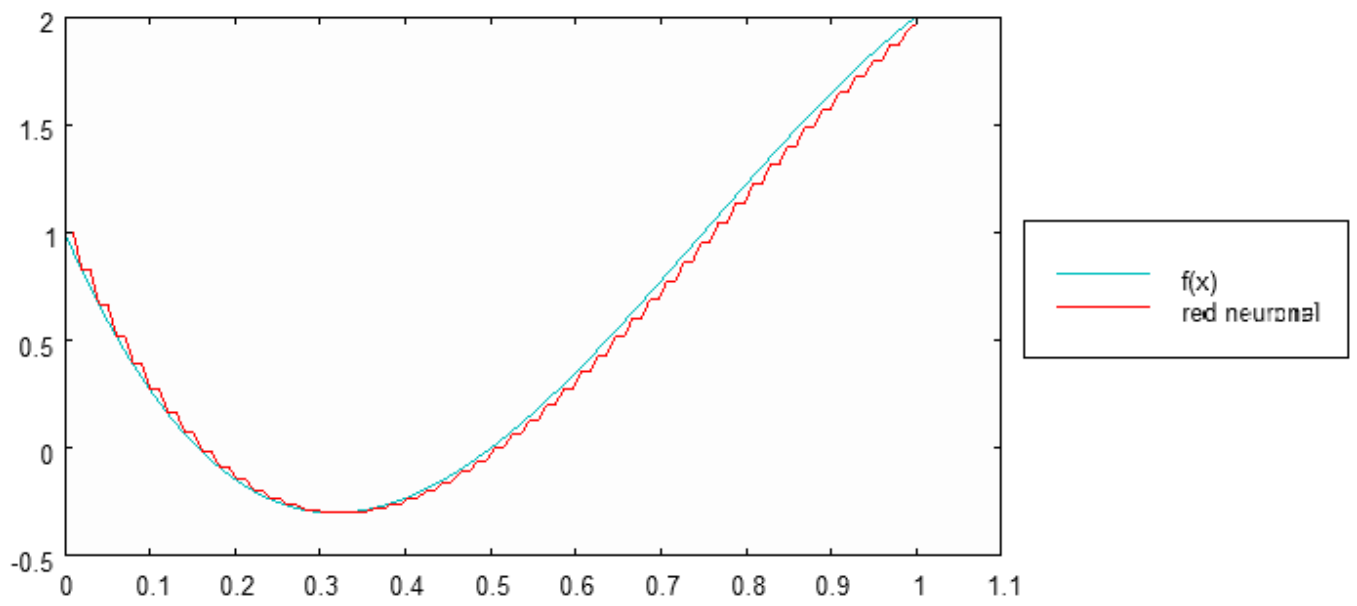


Figura 2.7: Gráfica de la aproximación a $f(x)$ para $n = 50$

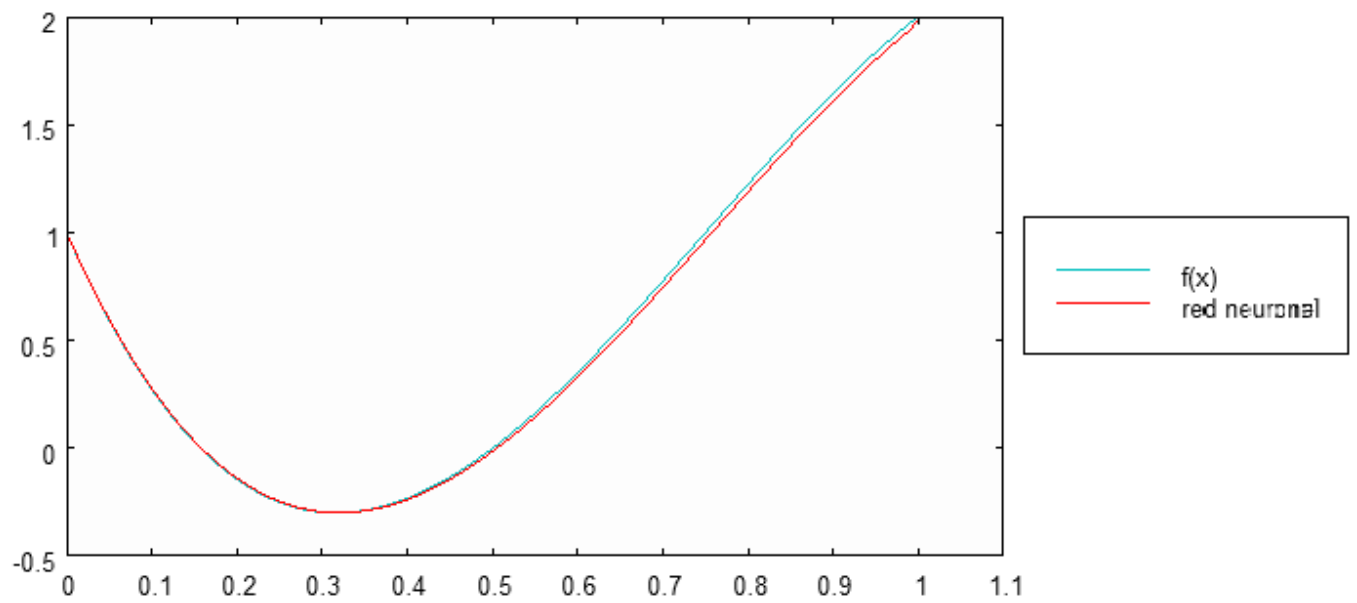


Figura 2.8: Gráfica de la aproximación a $f(x)$ para $n = 100$

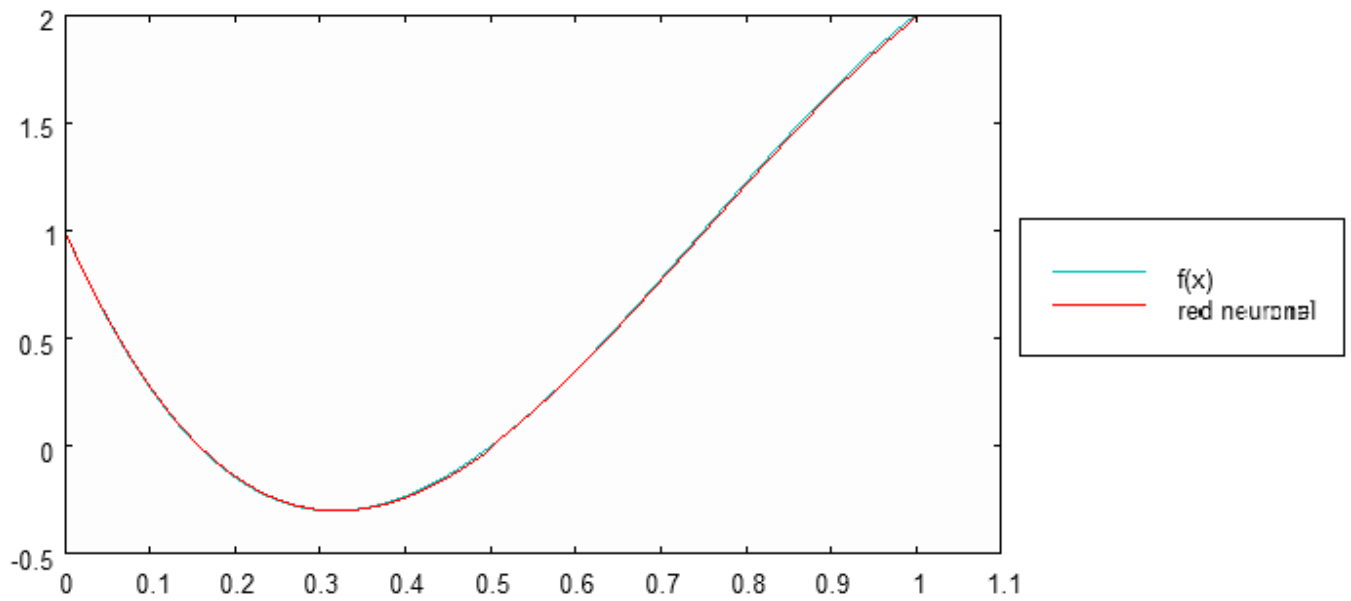


Figura 2.9: Gráfica de la aproximación a $f(x)$ para $n = 200$

La siguiente tabla de resultados muestra cómo el error máximo entre $f(x)$ y la red neuronal va decreciendo. Este error máximo es la norma $\|f - q\|$ calculada de la forma aproximada, ya que $E_{3n,\sigma}^{\Phi}(f) \leq \|f - g\|$:

n	Error máximo
15	0.5134
25	0.2947
50	0.1316
100	0.0448
200	0.0182

Cuadro 2.1: Errores máximos para los valores de n

Capítulo 3

Conclusiones

Es bien sabido que las redes neuronales son una herramienta muy potente para aproximar funciones continuas. 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 $[0,1]$ y el de las redes neuronales activadas mediante funciones sigmoideas. 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 sigmoideas de activación. Finalmente, un ejemplo numérico corrobora los resultados teóricos.

3.1. Ámbito formativo

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

Muchas de las técnicas de análisis matemático, empleadas en el estudio de los teoremas, las he ido desarrollando a lo largo de estos años y me han permitido entender mejor el comportamiento teórico de las redes neuronales, así como poder expresar en lenguaje matemático el desarrollo de todos los razonamientos.

Este trabajo me ha permitido familiarizarme con la herramienta de procesador de textos LaTeX, cuyo uso es habitual en los textos científicos. Otro reto ha sido la utilización del programa FreeMat, lo cual me ha permitido refrescar mis conocimientos de MatLab. Tengo que reconocer,

que los videotutoriales de YouTube me han resultado de gran ayuda.

Para finalizar, considero que ha sido muy útil la ayuda recibida por parte de mis tutores, ya que han sabido guiarme desde el principio y me han ayudado cuando me he estancado en algún punto.

3.2. Ámbito personal

En el ámbito personal, este trabajo me ha servido para darme cuenta de todo lo que he aprendido a lo largo del grado y de que soy capaz de enfrentarme a nuevos retos gracias a la formación obtenida. En general, he crecido como persona, me siento más capaz de cumplir proyectos y la confianza en mi misma ha aumentado viendo cómo he podido resolver y entender todos los resultados teóricos y el ejemplo numérico que se han presentado durante el desarrollo de este proyecto.

También me ha ofrecido la oportunidad de conocer el tema de redes neuronales, el cual hasta ahora era casi desconocido para mi y despertó mi curiosidad desde el principio.

En general, estoy muy satisfecha del resultado final de esta memoria y del desarrollo personal que he adquirido con su realización.

Bibliografía

- [1] G. CYBENKO, *Approximation by Superpositions of a Sigmoidal Function*. Mathematics of Control, Signals, and Systems, **2**, (1989), 303-314.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] B. L. KALMAN AND S. C. KWASNY,, *Why tanh: choosing a sigmoidal function*. IJCNN International Joint Conference on Neural Networks **4**, (1992), 578- 581.
- [6] 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.
- [7] M. V. MEDVEDEVA, *On sigmoidal functions*. Moscow University Mathematics Bulletin, **53**, (1998), no. 1, 16-19.