Grado en Matemática Computacional

Proyecto Final de Grado

Clasificación mediante redes neuronales

Autor:
Alba Hervías Francisco

Tutores académicos:
Juan José Font Ferrandis
Sergio Macario Vives

Fecha de lectura: Julio de 2023
Curso académico 2022/2023

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 cinco capítulos. En el primero de ellos se introducen el tema a tratar y los objetivos que persigue este proyecto. A continuación, en el segundo capítulo, se presentan todas las definiciones o lemas necesarios. El tercer capítulo se centra en el Teorema de Extensión de Tietze y en varios conceptos necesarios para poder demostrarlo. El cuarto capítulo contiene los resultados principales relacionados con la Clasificación, y las redes neuronales. Finalmente, en el último capítulo, se presentan las conclusiones.

Palabras clave

Clasificación, Teorema de Extensión de Tietze, funciones continuas, redes neuronales.

Keywords

Classification, Tietze Extension Theorem, continuous functions, neural networks.

Índice general

Capítulo 1

Introducción

Desde la antigüedad hemos querido clasificar objetos o fenómenos en categorías. Por ejemplo, los antiguos egipcios ya clasificaban plantas y animales en jeroglíficos. En el siglo XVIII, Carl Lineo desarrolló la taxonomía y más tarde, en el siglo XIX, Dmitri Mendeléyev la tabla periódica, ambos sistemas de clasificación. Con la aparición de la informática en el siglo XX, se empezaron a desarrollar las primeras máquinas de clasificación automática, pero ha sido en las últimas décadas cuando se han desarrollado algoritmos y técnicas que permiten a las computadoras aprender a clasificar todo aquello que se desee: señales de tráfico, expresiones faciales e incluso texturas de imágenes, sin necesidad de programación explícita.

Desde un punto de vista matemático, la idea es, partiendo de m conjuntos disjuntos, $K_1, ..., K_m$, de aquello que se quiera clasificar (por ejemplo, señales de tráfico), conseguir asignar a cada subconjunto K_j un valor real C_j. En nuestro contexto asumiremos que K_j son subconjuntos compactos de un espacio normado para facilitar la clasificación mediante el uso de técnicas matemáticas.

Para realizar este trabajo, desde el ámbito de las matemáticas, se han desarrollado las estructuras, técnicas o algoritmos necesarios para clasificar elementos dentro de grupos. Siguiendo con el ejemplo de las señales de tráfico, todas las señales tienen características visuales distintivas, colores, formas y símbolos. Las estructuras de Clasificación pueden analizar estas características y asignar cada señal a un grupo (de prohibición, de obligación, advertencias, entre otras).

En la actualidad hay un conjunto muy amplio de estructuras de Clasificación:

- Árboles de decisión:
  Se basan en una serie de preguntas binarias que dividen los datos en subconjuntos más pequeños, de modo que cada nodo representa una pregunta y cada rama una posible

respuesta.

- Regresión logística:
  Clasificador binario que utiliza una función logística para obtener una predicción entre 0 y 1 donde el valor de dicha predicción representa las probabilidades de que cada elemento pertenezca a una categoría u otra.

- K-vecinos más próximos:
  Los nuevos puntos se van clasificando en función de su proximidad con los puntos de entrenamiento previo.

- Redes neuronales.

A la hora de construir estos clasificadores se han utilizado tanto métodos estadísticos, en el caso de la regresión logística, como técnicas de aprendizaje automático, en clasificadores como redes neuronales y árboles de decisión, entre otros. La elección de cada estructura depende de muchos factores: el tipo de datos que se quieran clasificar, el problema que se quiera resolver o los objetivos finales de la clasificación.

De entre todas las estructuras, nos centraremos en las redes neuronales, una de las más utilizadas en el mundo del aprendizaje automático e inteligencia artificial. En esencia, su funcionamiento es similar al de las neuronas dentro del cerebro humano. Cada red neuronal está formada por varios nodos (o neuronas artificiales) interconectados. Cada neurona tiene múltiples señales de entrada, cada una con un peso asignado y un valor asociado a la entrada. Con toda esta información la neurona artificial realiza una suma ponderada y sobre dicha suma se ejecuta la función de activación, que es la que determinará la salida que tendrá la red neuronal. Esta viene definida 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*. La utilización de una función de activación u otra depende de cuál sea el objetivo de la red neuronal.

Es importante resaltar que, aunque a lo largo de este proyecto asociaremos redes neuronales con Clasificación, este no es el principal objetivo de las redes neuronales, sino la predicción numérica, donde tanto los elementos de entrada como los de salida son valores numéricos. Aún así, las redes neuronales se han podido adaptar a la Clasificación mediante el uso de métodos de representación de datos que transforman los datos originales en una representación numérica que sí pueda ser interpretada por la red neuronal. Aún no siendo el objetivo principal, la clasificación mediante redes neuronales presenta muchas ventajas, entre ellas, el hecho de que son aproximadores universales, es decir, pueden aproximar cualquier función continua con una precisión más que aceptable.

La motivación principal de este proyecto es, por tanto, mostrar el uso de las redes neuronales para abordar problemas de Clasificación.

Como se ha comentado previamente, este trabajo se ha dividido en cinco capítulos. En el primero de ellos se introducen el tema a tratar y los objetivos que persigue este proyecto. A continuación, en el segundo capítulo, se presentan todas las definiciones o lemas necesarios. El tercer capítulo se centra en el Teorema de Extensión de Tietze y en varios conceptos necesarios para poder demostrarlo. El cuarto capítulo contiene los resultados principales relacionados con la Clasificación, y las redes neuronales. Finalmente, en el último capítulo, se presentan las conclusiones.

Capítulo 2

Resultados preliminares

En esta parte se van a presentar una serie de definiciones necesarias para el estudio que se va a realizar.

Definición 1. *Sean X e Y espacios topológicos y $f_n : X \longrightarrow Y$ una sucesión de funciones. Se dice que la sucesión $\{f_n\}$ converge puntualmente a la función $f : X \longrightarrow Y$ si para todo $x \in X$ y $\epsilon > 0$ existe un $n_0 \in \mathbb{N}$ tal que $|f_n(x) - f(x)| < \epsilon$ para $n \geq n_0$.*

Definición 2. *Sea X un espacio topológico, (Y,d) un espacio métrico y $f_n : X \longrightarrow Y$ una sucesión de funciones. La sucesión $\{f_n\}$ converge uniformemente a $f : X \longrightarrow Y$ si para cualquier $\epsilon > 0$, existe $N > 0$ tal que $d(f(x), f_n(x)) < \epsilon$ para cualquier $x \in X$ y para cualquier $n > N$.*

Definición 3. *Un espacio topológico (X, T) es normal si y sólo si para cada par de conjuntos cerrados disjuntos $A \subset X$ y $B \subset X$, existe una función continua $f : X \longrightarrow [a, b]$ tal que $f(A) = a$ y $f(B) = b$.*

Lema 1 (Lema de Urysohn)**.** *Si X es un espacio normal, entonces para cada par de subconjuntos cerrados disjuntos, A y B, de X, existe una función continua $f : X \longrightarrow [0, 1]$ tal que $f(A) \subseteq \{0\}$ y $f(B) \subseteq \{1\}$.*

Definición 4. *Un homeomorfismo entre dos espacios topológicos es una aplicación continua biyectiva, cuya inversa es continua.*

Teorema 1 (Teorema de homeomorfismo)**.** *El conjunto de números reales $\mathbb{R}$ es homeomorfo al intervalo abierto $(-1, 1)$.*

Demostración. Definimos h como:

$$h : \mathbb{R} \longrightarrow (-1, 1)$$

$$x \longrightarrow tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

Y vamos a demostrar que es un homeomorfismo Primero veamos que h es inyectiva, suponemos que $tanh(x_1) = tanh(x_2)$ para $x_1, x_2 \in \mathbb{R}$

$$\begin{aligned}
\frac{e^{x_1} - e^{-x_1}}{e^{x_1} + e^{-x_1}} &= \frac{e^{x_2} - e^{-x_2}}{e^{x_2} + e^{-x_2}} \\
(e^{x_1} - e^{-x_1})(e^{x_2} + e^{-x_2}) &= (e^{x_1} + e^{-x_1})(e^{x_2} - e^{-x_2}) \\
e^{x_1 - x_2} &= e^{-x_1 + x_2} \\
x_1 - x_2 &= -x_1 + x_2 \\
x_1 &= x_2
\end{aligned}$$

Luego $h(x)$ es inyectiva. Además, es sobreyectiva, ya que dado $y \in (-1, 1)$, buscamos $x \in \mathbb{R}$ tal que $h(x) = y$. Despejando,

$$\begin{aligned}
\frac{e^x - e^{-x}}{e^x + e^{-x}} &= y \\
\frac{e^x(1 - e^{-2x})}{e^x(1 + e^{-2x}} &= y \\
1 - e^{-2x} &= y + ye^{-2x} \\
1 - y &= (1 + y)e^{-2x} \\
e^{-2x} &= \frac{1 - y}{1 + y} \\
e^{2x} &= \frac{1 + y}{1 - y} \\
x &= \frac{1}{2} \log \frac{1 + y}{1 - y}
\end{aligned}$$

Así concluimos que la función es biyectiva. Además $h(x)$ es continua por ser composición de funciones continuas y su función inversa $h^{-1}(y) = \frac{1}{2} \log \frac{1+y}{1-y}$ es continua en todo su dominio $(-1, 1)$. Luego h es un homeomorfismo por ser una función biyectiva continua con inversa continua. $\square$

Necesitaremos también algunos resultados sobre convergencia uniforme.

Lema 2. *Sea X un espacio topológico e (Y, d) un espacio métrico. Asumimos que $f_n : X \longrightarrow Y$ es una sucesión de funciones que converge uniformemente a $f : X \longrightarrow Y$. Si cada f_n es continua, entonces f es continua.*

Demostración. Sea $U \subseteq Y$ un conjunto abierto. Para demostrar que f es continua queremos ver que $f^{-1}(U)$ es abierto. Sea $x_0 \in f^{-1}(U)$, es decir, $f(x_0) \in U$.

Por ser U abierto, existe $\epsilon > 0$ tal que la bola abierta $B_\epsilon(f(x_0)) \subseteq U$. Por la convergencia uniforme podemos tomar $N > 0$ tal que $d(f(x), f_n(x)) < \frac{\epsilon}{3}$ para todo $x \in X$. Definimos $V = f_N^{-1}(B_{\frac{\epsilon}{3}}(f_N(x_0)))$. Como f_N es continua, V es un entorno abierto de x_0. Si $x \in V$ entonces $f_N(x) \in B_{\frac{\epsilon}{3}}(f_N(x_0))$ y, por tanto,

$$d(f(x), f(x_0)) < d(f(x), f_N(x)) + d(f_N(x), f_N(x_0)) + d(f_N(x_0)), f(x_0)) < \frac{\epsilon}{3} + \frac{\varepsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

Luego, si $x \in V \quad d(f(x), f(x_0)) < \epsilon$, y en consecuencia, $f(x) \in B_\epsilon(f(x_0)) \subset U$. Es decir, $x_0 \in V \subseteq f^{-1}(U)$, lo que significa que f es continua. □

Teorema 2 (Criterio de Weierstrass)**.** *Sea K un espacio compacto. Una condición suficiente para que la serie $\sum_{n=1}^{\infty} f_n(t)$ de funciones $f_n : K \longrightarrow \mathbb{R}$ sea uniformemente convergente es que exista una serie numérica convergente $\sum_{n=1}^{\infty} \rho_n$ verificando $|f_n(t)| < \rho_n$ para todo $t \in K$ y todo $n \in \mathbb{N}$.*

Capítulo 3

Teorema de Tietze

En este capítulo se demuestra el Teorema de Extensión de Tietze que va a ser una herramienta básica para relacionar el problema de Clasificación con las redes neuronales. En todo momento, se considerará que estamos trabajando con una red neuronal como la definida en la Introducción, es decir,

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

Lema 3. *Sea X un espacio topológico normal , $A \subset X$ un subespacio cerrado y $f : A \longrightarrow \mathbb{R}$ una función continua acotada, es decir, para algún $C > 0$, $|f(x)| < C$ para todo $x \in A$. Entonces existe una función continua $g : X \longrightarrow \mathbb{R}$ tal que $|g(x)| < \frac{1}{3}C$ para cualquier $x \in X$ y $|f(x) - g(x)| < \frac{2}{3}C$ para cualquier $x \in A$.*

Demostración. Definimos:

$$Y = f^{-1}\left(\left[-C, -\frac{1}{3}C\right]\right)$$
$$Z = f^{-1}\left(\left[\frac{1}{3}C, C\right]\right)$$

Puesto que f es continua, Y y Z son conjuntos cerrados en A y al ser este cerrado en X, implica que Y, Z son cerrados en X. Como además, $Y \cap Z = \emptyset$, por el Lema 1, existe una función continua $h : X \longrightarrow [0, 1]$ tal que $h(Y) = 0$ y $h(Z) = 1$.

Finalmente, definimos

$$g(x) = \frac{2C}{3}\left(h(x) - \frac{1}{2}\right).$$

Veamos ahora si g cumple los requisitos de enunciado;

$$|g(x)| = \left|\frac{2C}{3}(h(x) - \frac{1}{2})\right| = \frac{2C}{3}\left|h(x) - \frac{1}{2}\right| < \frac{2C}{3}\frac{1}{2} = \frac{C}{3}$$

Sea $x \in A$, vamos a estudiar el valor de $|f(x) - g(x)|$ distinguiendo tres posibilidades:

- $f(x) \in \left[-C, -\frac{1}{3}C\right]$, es decir, $x \in Y$. Entonces, $h(x) = 0$ y

$$|f(x) - g(x)| = \left|f(x) - \frac{2C}{3}\left(h(x) - \frac{1}{2}\right)\right| = \left|f(x) + \frac{1}{3}C\right| < \frac{2}{3}C$$

- $f(x) \in \left(-\frac{1}{3}C, \frac{1}{3}C\right)$. Entonces $|f(x)| < \frac{1}{3}C$ y $|g(x)| < \frac{1}{3}C$, luego

$$|f(x) - g(x)| \leqslant |f(x)| + |g(x)| < \frac{1}{3}C + \frac{1}{3}C = \frac{2}{3}C$$

- $f(x) \in \left[\frac{1}{3}C, C\right]$. Entonces $x \in Z$ y $h(x) = 1$, por lo que

$$|f(x) - g(x)| = \left|f(x) - \frac{2C}{3}\left(h(x) - \frac{1}{2}\right)\right| = \left|f(x) - \frac{1}{3}C\right| < \frac{2}{3}C$$

Luego $|f(x) - g(x)| < \frac{2}{3}C$ para todo $x \in A$. □

Teorema 3 (Teorema de extensión de Tietze)**.** *Sea X un espacio normal, $A \subseteq X$ un subespacio cerrado y $f : A \longrightarrow [a, b]$ una función continua para $[a, b] \subseteq \mathbb{R}$. Entonces existe una función continua $\overline{f} : X \longrightarrow [a, b]$ de forma que $\overline{f}|_A = f$.*

Demostración. Sin pérdida de generalidad, tomaremos $[a, b] = [0, 1]$. Vamos a construir funciones continuas $g_n : X \longrightarrow \mathbb{R}$ para $n = 1, 2, ...$ que cumplan:

$$|g_n(x)| < \frac{1}{3}\left(\frac{2}{3}\right)^{n-1} \quad \forall x \in X \tag{3.1}$$

$$\left|f(x) - \sum_{i=1}^{n} g_i(x)\right| < \left(\frac{2}{3}\right)^n \quad \forall x \in A \tag{3.2}$$

Lo demostraremos por inducción. Puesto que $|f(x)| < 1$, f está acotada. Por el Lema 3, existe $g_1 : X \longrightarrow \mathbb{R}$ tal que $|g_1(x)| < \frac{1}{3}$ para todo $x \in X$ y $|f(x) - g_1(x)| < \frac{2}{3}$ para todo $x \in A$. Suponemos que $g_1, g_2, ..., g_n$ cumplen las condiciones 3.1 y 3.2. Aplicando el Lema 3 sobre $f - \sum_{i=1}^{n} g_i$ y tomando $C = \left(\frac{2}{3}\right)^n$, encontramos g_{n+1} tal que cumple las condiciones,

$$|g_{n+1}(x)| < \frac{1}{3}\left(\frac{2}{3}\right)^n$$

$$\left| f - \sum_{i=1}^{n} g_i - g_{n+1} \right| < \frac{2}{3}\left(\frac{2}{3}\right)^n$$

Sean $\overline{f_n} = \sum_{i=1}^{n} g_n$ y $\overline{f} = \sum_{i=1}^{\infty} g_n$. Puesto que, $|g_i(x)| < \frac{1}{3}\left(\frac{2}{3}\right)^i$ y $\sum_{i=1}^{\infty} \frac{1}{3}\left(\frac{2}{3}^i\right)$ es convergente, por el Teorema 2 $\overline{f_n}(x) = \lim_{x\to\infty} \overline{f}(x)$ está bien definida y la sucesión $f_n(x)$ converge uniformemente a $\overline{f}$, de donde se deduce que se cumple 3.1 y

$$\left| f(x) - \sum_{i=1}^{n+1} g(x) \right| < \left(\frac{2}{3}\right)^{n+1}$$

Luego la condición 3.2 también se cumple, y así queda demostrada la construcción de las funciones g_n por inducción.

Puesto que g_n son funciones continuas, entonces $\overline{f_n}$ también son funciones continuas y por el Lema 2, $\overline{f}$ es continua.
Por 3.2 sabemos que:

$$\left| f - \sum_{i=1}^{n} g_i \right| < \left(\frac{2}{3}\right)^n \quad \forall x \in A,$$

$$|f - \overline{f_n}| < \left(\frac{2}{3}\right)^n \quad \forall x \in A.$$

Tomando límites cuando n tiende a infinito se concluye que $f(x) = \overline{f(x)}$ para cualquier $x \in A$. $\square$

Teorema 4 (Teorema de extensión de Tietze)**.** *Sea X un espacio topológico normal , A un subespacio cerrado de X y $f : A \longrightarrow \mathbb{R}$ una función continua. Entonces, existe $\overline{f} : X \longrightarrow \mathbb{R}$ continua, de forma que $\overline{f}(x) = f(x) \quad \forall x \in A$.*

Demostración. En primer lugar demostraremos que para una función $g : A \longrightarrow (-1, 1)$ podemos encontrar una función continua $\overline{g} : X \longrightarrow (-1, 1)$ de forma que $\overline{g}|_A = g$. Una vez demostrado este caso utilizaremos el homeomorfismo $h : \mathbb{R} \longrightarrow (-1, 1)$, obtenido por el Teorema 1, para demostrar el caso general.

Asumimos que $g : A \longrightarrow (-1,1)$ es una función continua, por el Teorema 3, sabemos que existe $g_1 : X \longrightarrow [-1,1]$ tal que $g_1|_A = g$. Tomamos $B := g_1^{-1}(\{-1,1\})$. B es cerrado en X y $A \cap B = \emptyset$, ya que, si $x \in A \cap B$ entonces $g_1(x) \in \{-1,1\}$, y al ser $x \in A$, $g_1(x) \in (-1,1)$, lo cual es una contradicción.

Por el Lema 1, existe una función continua $k : X \longrightarrow [0,1]$ de forma que $k(x) = 0$ si $x \in B$ y $k(x) = 1$ si $x \in A$. Definimos $\overline{g}(x) = k(x) \cdot g_1(x) : X \longrightarrow [-1,1] \subseteq \mathbb{R}$ que es una función continua por ser producto de funciones continuas. Además se cumple que:

- si $g_1(x) \in \{-1,1\}$ entonces $x \in B$, por lo que $\overline{g}(x) = 0 \cdot g_1(x) = 0$
- si $g_1(x) \in (-1,1)$ sabemos que $|\overline{g}| = |k \cdot g_1| = |k| \cdot |g_1| < 1 \cdot 1 = 1$ luego $\overline{g}(x) \in (-1,1)$

Por tanto, $\overline{g}(x)$ no puede alcanzar los valores 1 ni -1. Es decir, $\overline{g} : X \longrightarrow (-1,1)$. Por otra parte, si $x \in A$, entonces $\overline{g}(x) = k(x) \cdot g_1(x) = 1 \cdot g_1(x) = g_1(x) = g(x)$.

Finalmente, componiendo el homeomorfismo h sobre las funciones g_1 y g, se tiene que $\overline{f} = \overline{h} g_1$ y $f = h^{-1} g$ y se cumple

$$f = h^{-1} g = h^{-1} g_1 = h^{-1} h \overline{f} = \overline{f} \quad \forall x \in A.$$

Luego $\overline{f}(x) = f(x)$ para todo $x \in A$. $\square$

Capítulo 4

Clasificación

En este capítulo abordamos el problema de Clasificación mediante redes neuronales aprovechando que el Teorema de Extensión de Tietze nos proporciona un clasificador.

Definición 5. *Entendemos por problemas de Clasificación aquellos en los que el objetivo es asignar una etiqueta (valor numérico) a cada uno de los datos del conjunto de entrada.*

El método de clasificación que usaremos en este trabajo será el de encontrar una función continua que asigne valores numéricos diferentes a cada objeto matemático que se quiera clasificar. Como las redes neuronales pueden aproximar funciones continuas, se pueden utilizar para aproximar la función de clasificación. Sin embargo, es esperable que su salida no coincida con absoluta precisión con una de las etiquetas numéricas.

Definición 6. *Dados intervalos disjuntos, cada uno de los cuales contiene una de las etiquetas numéricas que se quieren asignar en un problema de Clasificación, entenderemos que una red neuronal clasifica los objetos cuando, para cada dato de entrada, proporciona una salida que pertenece a alguno de dichos intervalos.*

Teorema 5. *Sean N conjuntos cerrados disjuntos $K_1, K_2, ..., K_N \subset \mathbb{R}^n$, N valores reales arbitrarios $C_1, C_2, ..., C_N$ y un conjunto cerrado arbitrario X disjunto de cualquier conjunto K_i para $i = 1, ..., N$. Entonces, existe una función continua $f(x)$ tal que $f(x) = C_i$ para $x \in K_i$ y $f(x) = C_0$ para $x \in X$, donde C_0 es un valor real distinto de $C_1, C_2, ...C_N$.*

Demostración. Construimos la función

$$f(x) = \begin{cases} C_i & \text{si} \quad x \in K_i,\ i = 1, ..., N \\ \\ C_0 & \text{si} \quad x \in X \end{cases}$$

El conjunto $A = K_1 \cup K_2 \cup ... \cup K_N \cup X$ es un conjunto cerrado por ser unión finita de conjuntos cerrados. La función f es continua en cada cerrado por ser constante y al ser los cerrados disjuntos, la función f es continua en A. Por el Teorema de Tietze, existe $\overline{f} : \mathbb{R}^n \longrightarrow \mathbb{R}$ continua, tal que $f(x) = \overline{f}(x)$ para cualquier $x \in A$. $\square$

Definición 7. *Definimos* $\overline{C}(\mathbb{R}^n) = \left\{ f(x) \in C(R^n) : \textit{existe } \lim_{\|x\| \to +\infty} f(x) \right\}$

Teorema 6. *Cualquier familia de N conjuntos compactos disjuntos de $\mathbb{R}^n$ puede ser clasificada mediante una red neuronal de la forma*

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

siempre que estas redes sean densas en el espacio de funciones continuas $\overline{C}(\mathbb{R}^n)$.

Demostración. Sean N conjuntos compactos disjuntos $K_i \subseteq \mathbb{R}^n$, $i = 1, ..., N$, y N valores reales distintos $C_1, C_2, ..., C_N$. Por ser $K_1 \cup ... \cup K_N$ un conjunto compacto, existe una bola abierta que lo contiene, es decir, existe $r > 0$ tal que

$$B_r(0) \supseteq K_1 \cup ... \cup K_N.$$

Sea $X = \mathbb{R}^n \setminus B_{2r}(0)$ un cerrado y sea $C_0 \in \mathbb{R}$ distinto de cada uno de los valores C_i, $i = 1, ..., N$. Entonces aplicando el Teorema 5 existe una función continua $f \in C(\mathbb{R}^n)$ tal que $f(x) = C_i$, $x \in K_i$, $i = 1, ..., N$. Además $f(x) = C_0$ si $x \in X$, lo que implica que $\lim_{\|x\| \to \infty} f(x) = C_0$. Por tanto, $f(x) \in \overline{C}(\mathbb{R}^n)$.

Además, como las redes son densas en $\overline{C}(\mathbb{R}^n)$, dado $\epsilon > 0$, existe una red neuronal $N(x)$ tal que

$$|N(x) - f(x)| < \epsilon, \quad \text{para todo } x \in X.$$

Tomamos

$$\epsilon = \frac{1}{2} \min_{i,j=1,\dots,N} \{|C_i - C_j|\}.$$

Dado $x \in K_i$,

$$|N(x) - f(x)| = |N(x) - C_i| < \epsilon.$$

Entonces,

$$N(x) \in [C_i - \epsilon, C_i + \epsilon]$$

y, puesto que los intervalos son disjuntos entre sí, esto permite la clasificación. $\square$

A continuación, mostraremos algun tipo de red neuronal densa en $\overline{C}(\mathbb{R})$. En primer lugar necesitamos algunos conceptos básicos.

Definición 8. *Una función sigmoidal es una aplicación*

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

que cumple $\lim\limits_{t \to -\infty} \sigma(t) = 0$ *y* $\lim\limits_{t \to +\infty} \sigma(t) = 1$.

Definición 9. *Dada una función acotada* $\sigma : \mathbb{R} \to \mathbb{R}$*, definimos la norma* $||\sigma||_\infty$ *como*

$$||\sigma||_\infty := \sup_{x \in \mathbb{R}} |\sigma(x)|$$

Lema 4. *Si* f *es una función continua en* $\mathbb{R}$ *cumpliendo que* $\lim\limits_{|x| \to \infty} f(x) = 0$*, entonces* f *es una función uniformemente continua en* $\mathbb{R}$.

Demostración. Sea $\epsilon > 0$. Debemos encontrar $\delta > 0$ tal que $|x - y| < \delta$ implique que $|f(x) - f(y)| < \epsilon$.

Dado que $\lim_{|x| \to \infty} f(x) = 0$, existe $M > 0$ tal que si $|x| \geq M$, entonces $|f(x)| < \frac{\epsilon}{3}$.

Puesto que f es continua en el intervalo compacto $[-M, M]$, entonces es uniformemente continua y, por tanto, existe $\delta > 0$ tal que si $x, y \in [-M, M]$ y $|x - y| < \delta$, entonces $|f(x) - f(y)| < \frac{\epsilon}{3}$.

Así, dados $x, y \in \mathbb{R}$, con $|x - y| < \delta$, distinguiremos los siguientes tres casos:

1. Si $x, y \in [-M, M]$, es evidente que se cumple que

$$|f(x) - f(y)| < \epsilon/3 < \epsilon.$$

2. Si $|x|$, $|y| > M$, entonces se cumple que $|f(x)| < \epsilon/3$ y $|f(y)| < \epsilon/3$, y por tanto,

$$|f(x) - f(y)| < 2\epsilon/3 < \epsilon.$$

3. Si $x \in [-M, M]$ y $|y| > M$. Entonces, $|f(y)| < \epsilon/3$ y

$$|f(x) - f(y)| \leq |f(x) - f(M)| + |f(M) - f(y)| \leq |f(x) - f(M)| + |f(M)| + |f(y)| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

$\square$

Definición 10. *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 11. *Definimos la familia de funciones $\psi_{n,\sigma}$ como*

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

Definición 12. *Definimos el error de aproximación entre una función continua f y $\psi_{n,\sigma}$ como*

$$E^{\psi}_{n,\sigma}(f) := \inf_{g \in \psi_{n,\sigma}} ||f - g||_\infty$$

Teorema 7. *Sea $\sigma(x)$ una función sigmoidal acotada y sea $f(x) \in \overline{C}(\mathbb{R})$, con $\lim_{x \to \infty} f(x) = a$ y $\lim_{x \to -\infty} f(x) = b$ donde $a, b \in \mathbb{R}$. Entonces para cada n entero positivo,*

$$E^{\Psi}_{n,\sigma} \leq (||\sigma|| + 1)\omega(f, \frac{2A_1}{n+1})$$

donde la constante A_1 depende de f y σ.

Demostración. Dado que f es uniformemente continua en $\mathbb{R}$, existe $M > 0$ tal que $|f(x)| \leq M$ para cualquier $x \in \mathbb{R}$. Luego, para cualquier $\beta > 0$

$$\omega(f, \beta) = sup\{|f(x) - f(y)| : |x - y| \leq \beta\} \leq 2M.$$

Dado $\epsilon > 0$, fijamos un número natural n, y definimos,

$$\gamma = \frac{\epsilon}{2M(n+1)}.$$

Por la definición de función sigmoidal, existe una constante $A_1 \geq 1$, tal que para cualquier $x \geq A_1$

$$|f(x) - a| < \frac{\epsilon}{2}, \quad |\sigma(x) - 1| < \gamma \tag{4.1}$$

y, para cualquier $x \leq -A_1$,

$$|f(x) - b| < \frac{\epsilon}{2}, \quad |\sigma(x)| < \gamma. \tag{4.2}$$

Sea α_1 un número arbitrario tal que $0 < \alpha_1 < \frac{A_1}{2(n+1)}$. Entonces:

$$\left|\sigma\left(\frac{A_1}{\alpha_1}x\right) - 1\right| < \gamma \quad \forall x \geq \alpha_1,$$

$$\left|\sigma\left(\frac{A_1}{\alpha_1}x\right)\right| < \gamma \quad \forall x \leq -\alpha_1.$$

Sea $x_j = -A_1 + 2A_1(\frac{j}{n+1})$ para $j = 0, ..., n, n+1$. Entonces definimos

$$g(x) = f(x_1) + \sum_{i=1}^{n}(f(x_{i+1} - f(x_i))\sigma\left(\frac{A_1}{\alpha_1}(x - x_1)\right).$$

Queremos demostrar que $|f(x) - g(x) \leq (||\sigma|| + 1)\omega\left(f, \frac{2A_1}{(n+1)}\right)$ para cualquier $x \in \mathbb{R}$. Distinguimos cuatro situaciones distintas:

1. $x \in [x_0, x_0 + \alpha_1] \cup (\bigcup_{i=1}^{n}[x_k - \alpha_1, x_k + \alpha_1]) \cup [x_{n+1} - \alpha_1, x_{n+1}]$
2. $x \in \bigcup_{k=0}^{n}(x_k + \alpha_1, x_{k+1} - \alpha_1)$
3. $x \in (x_{n+1}, \infty)$
4. $x \in (-\infty, x_0)$

CASO 1. Suponemos que $x \in [x_0, x_0 + \alpha_1] \cup (\bigcup_{i=1}^{n}[x_k - \alpha_1, x_k + \alpha_1]) \cup [x_{n+1} - \alpha_1, x_{n+1}]$. Luego existe $1 \leq k \leq n$ tal que

- $x \in [x_k - \alpha_1, x_k + \alpha_1]$, luego:

$$x - x_i \leq \alpha_1 \text{ para } i = 0, ..., k-1$$

$$x - x_i \geq -\alpha_1 \text{ para } i = k+1, k+2, ..., n+1$$

- $x \in [x_{n+1} - \alpha_1, x_{n+1}]$ entonces $x - x_i \geq \alpha_1$ para $i = 0, 1, ..., n$
- $x \in [x_0, x_0 + \alpha_1]$ entonces $x - x_1 \leq -\alpha_1$ para $i = 1, 2, ..., n+1$

Entonces

$$
\begin{aligned}
g(x) - f(x) &= f(x_k) - f(x) + \sum_{i=1}^{k-1} (f(x_{i+1}) - f(x_i)) \left(\sigma \left(\frac{A_1}{\alpha_1}(x - x_i) \right) - 1 \right) \\
&\quad + (f(x_{k+1}) - f(x_k)) \sigma \left(\frac{A_1}{\alpha_1}(x - x_k) \right) + \sum_{i=k+1}^{n} (f(x_{i+1}) - f(x_i)) \left(\sigma \left(\frac{A_1}{\alpha_1}(x - x_i) \right) \right) \\
&\leq \omega(f, \alpha_1) + (k-1)\omega \left(f, \frac{2A_1}{n+1} \right) \gamma + ||\sigma|| \omega \left(f, \frac{2A_1}{n+1} \right) + (n-k)\omega \left(f, \frac{2A_1}{n+1} \right) \gamma \\
&= \omega(f, \alpha_1) + ||\sigma|| \omega \left(f, \frac{2A_1}{n+1} \right) + (n-1)\omega \left(f, \frac{2A_1}{n+1} \right) \gamma \\
&\leq \omega(f, \alpha_1) + ||\sigma|| \omega \left(f, \frac{2A_1}{n+1} \right) + \epsilon \leq \omega \left(f, \frac{2A_1}{n+1} \right) + ||\sigma|| \omega \left(f, \frac{2A_1}{n+1} \right) + \epsilon \\
&= (||\sigma|| + 1) \omega \left(f, \frac{2A_1}{n+1} \right) + \epsilon.
\end{aligned}
$$

CASO 2. Si $x \in \bigcup_{k=0}^{n} (x_k + \alpha_1, x_{k+1} - \alpha_1)$, entonces existe $0 \leq k \leq n$ tal que $x \in (x_k + \alpha_1, x_{k+1} - \alpha_1)$. Así

$$x - x_i > \alpha_1 \text{ para } i = 0, 1, ..., k \text{ y}$$

$$x - x_i < -\alpha_1 \text{ para } i = k+1, ..., k+n$$

Entonces

$$
\begin{aligned}
g(x) - f(x) &= f(x_{k+1}) - f(x) + \sum_{i=1}^{k} (f(x_{i+1}) - f(x_i)) \left(\sigma \left(\frac{A_1}{\alpha_1}(x - x_i) \right) - 1 \right) \\
&\quad + \sum_{i=k+1}^{n} (f(x_{i+1}) - f(x_i)) \left(\sigma \left(\frac{A_1}{\alpha_1}(x - x_i) \right) \right) \\
&\leq \omega(f, \alpha_1) + k\omega \left(f, \frac{2A_1}{n+1} \right) \gamma + (n-k)\omega \left(f, \frac{2A_1}{n+1} \right) \gamma \\
&\leq \omega \left(f, \frac{2A_1}{n+1} \right) + \epsilon \\
&\leq (||\sigma|| + 1) \omega \left(f, \frac{2A_1}{n+1} \right) + \epsilon.
\end{aligned}
$$

CASO 3. Si $x \in (x_{n+1}, \infty)$, entonces, para $k = 0, 1, ..., n$, se tiene que $x - x_k \geq \alpha_1$. Aplicando

(4.1),

$$
\begin{aligned}
g(x) - f(x) &= \sum_{i=1}^{n} (f(x_{i+1}) - f(x_i)) \left(\sigma \left(\frac{A_1}{\alpha_1}(x - x_i) \right) - 1 \right) + f(x_{n+1}) - f(x) \\
&\leq n\omega \left(f, \frac{2A_1}{n+1} \right) \gamma + \epsilon \\
&\leq 2\epsilon \leq (||\sigma|| + 1)\omega \left(f, \frac{2A_1}{n+1} \right) + 2\epsilon.
\end{aligned}
$$

CASO 4. Si $x \in (-\infty, x_0)$, entonces $k = 1, ..., n, n+1$, $x - x_k \geq -\alpha_1$, y siguiendo el mismo procedimiento que en el Caso 3, con (4.2),

$$
\begin{aligned}
g(x) - f(x) &\leq \omega \left(f, \frac{2A_1}{n+1} \right) + \epsilon \\
&\leq 2\epsilon \leq (||\sigma|| + 1)\omega \left(f, \frac{2A_1}{n+1} \right) + 2\epsilon.
\end{aligned}
$$

Luego, sabiendo que tanto ϵ como α_1 son valores arbitrarios, resulta

$$
E_{n,\sigma}^{\Psi} \leq (||\sigma|| + 1)\omega \left(f, \frac{2A_1}{n+1} \right).
$$

□

El teorema anterior garantiza que las redes neuronales que tienen la forma de funciones de la familia $\psi_{n,\sigma}$ son densas en $\overline{C}(\mathbb{R})$ y, por tanto, se pueden utilizar para resolver problemas de Clasificación.

Capítulo 5

Conclusiones

Tomando la Clasificación como hilo conductor, en este proyecto hemos comprobado como dos herramientas matemáticas aparentemente desconectadas, el Teorema de Tietze y la redes neuronales, pueden unirse para proporcionar resultados útiles en el ámbito de los problemas de Clasificación.

Tras comentar todas aquellas definiciones y conceptos necesarios, nos hemos centrado en el teorema de extensión de Tietze, que garantiza poder extender una función continua definida en un subconjunto cerrado a una función continua en todo el espacio. El teorema, nos permite obtener una función continua de clasificación para N conjuntos compactos disjuntos que posteriormente puede ser aproximada mediante una red neuronal adecuada.

Por otra parte, las redes neuronales han venido utilizándose desde los años 90 como excelentes aproximadores de funciones continuas en muchos contextos. En esta memoria hemos estudiado uno de esos resultados con todo detalle, para su posterior aplicación a los problemas de Clasificación. Una vez introducidos todos estos conceptos, hemos visto que, con el uso de funciones sigmoidales, es posible aproximar determinadas funciones en el espacio $\overline{C}(\mathbb{R})$.

La elaboración de este proyecto me ha permitido poner en práctica muchos de los conocimientos adquiridos en el Grado en Matemática Computacional, permitiéndome comenzar un proyecto matemático desde cero, partiendo de las definiciones más básicas hasta llegar a los teoremas principales que respaldaban la idea principal. Además he adquirido nuevas habilidades en el uso de LaTex que me han permitido producir un documento de calidad con una presentación precisa y profesional. Otro de los aspectos que me gustaría resaltar es la importancia de referenciar las fuentes utilizadas, pero mediante un método riguroso de citación de bibliografía.

En el ámbito personal, he reafirmado una idea que a lo largo de todo el grado ha ido cogiendo más peso, y es que, a través del esfuerzo se consiguen las cosas. El esfuerzo requiere compromiso, perseverancia y disciplina, y son todas estas cosas las que me han hecho llegar hasta aquí.

Está claro que el esfuerzo no implica éxito, pero sí asienta las bases sobre las que crecer hasta llegar a él.

Bibliografía

[1] B. I. Hong and N. Hahm, *Approximation Order to a Functions in* $\overline{C}(\mathbb{R})$ *by Superposition of a Sigmoidal Function.* Applied Mathematics Letters, **15**, (2002), 591-597.

[2] G.B Huang, Y.Q. Chen and H.A. Barbi, *Classification Ability of Single Hidden Feedforward Neural Networks.* IIEE Transactions Neural Networks, **71**, (2000), 799-801.

[3] J. Munkres, *Topology.* Pearson, (2014).

[4] I. W. Sandberg, *General Structures for Classification.* IIEE Transactions on Circuits ans Systems, **71**, (1994), 372-376.