

# Métodos de predicción no paramétricos: aplicación a series temporales

by

ANTONIO LOPEZ SERRANO

Una tesis presentada en conformidad con los requisitos para el Master en Economía Finanzas y Computación.

Universidad de Huelva y Universidad Internacional de Andalucía

uhu.es

un  
i Universidad  
Internacional  
de Andalucía  
A

Noviembre 2017

# Métodos de predicción no paramétricos: aplicación a series temporales

Antonio López Serrano

Máster en Economía, Finanzas y Computación

José Manuel Bravo Caro

Universidad de Huelva y Universidad Internacional de Andalucía

Noviembre 2017

## Abstract

Data prediction has been acquiring greater attention and interest in recent years, since technological advances have led to the development of new and increasingly sophisticated predictive models, which have shown a constant improvement in the decision-making process of multiple and very different areas.

In this paper, the problem of predicting data from a non-parametric approach is addressed, from the perspective of using a weighted sum of a set of previous data in order to predict future data. In our case, unlike other approaches proposed in the existing literature, the sum is weighted by some weights obtained as a result of the resolution of a convex optimization problem, which in turn incorporates an adjustment parameter that allows the improvement of the performance of the predictor.

To evaluate the performance of this prediction method in a real scenario, a practical application is included making use of a solar power generation database. In this way, it is demonstrated that with the proposed predictor, adequate results can be obtained in a validation scheme based on the comparison of our predictor with other prediction methods, reaffirming in this way that this prediction method can be an interesting alternative to the prediction techniques based on classical methods.

Therefore, the main objective of this research is to provide a new approach to data prediction based on non-parametric regression methods, which minimizes the prediction error as much as possible,

in order to provide a useful tool that can be used in any field to improve the decision-making process.

Gel classification: C53, C63, C61

Keywords: prediction models, nonparametric regression, optimization techniques

## Resumen

La predicción de datos ha venido adquiriendo una mayor atención e interés en los últimos años, puesto que los avances tecnológicos han propiciado el desarrollo de nuevos modelos predictivos cada vez más sofisticados, que han logrado demostrar una mejora constante en el proceso de toma de decisiones de múltiples y muy diferentes ámbitos.

En este trabajo, se aborda el problema de la predicción de datos desde un enfoque no paramétrico, cuya base se sustenta en utilizar una suma ponderada de un conjunto de datos para realizar predicciones de datos futuros. En nuestro caso, a diferencia de otros enfoques propuestos en la literatura existente, la suma se pondera mediante unos pesos que se obtienen como resultado de la resolución de un problema de optimización convexo, el cual a su vez incorpora un parámetro de ajuste que permite mejorar el rendimiento del predictor.

Para evaluar el desempeño de este método de predicción en un escenario real, se incluye una aplicación práctica haciendo uso de la base de datos de energía fotovoltaica. De esta manera, se demuestra que con el predictor propuesto se pueden obtener resultados adecuados en un esquema de validación basado en la comparación de nuestro predictor con otros métodos de predicción, reafirmando de esta manera que este método de predicción puede ser interesante y alternativo a las técnicas de predicción basadas en métodos clásicos.

Por lo tanto, el objetivo principal de esta investigación es proporcionar un nuevo enfoque de predicción de datos basado en métodos de regresión no paramétrica, que logre minimizar el error de predicción tanto como sea posible, con el fin de aportar una herramienta útil que pueda ser utilizada en cualquier ámbito para la mejora del proceso de toma de decisiones.

Clasificación JEL: C53, C63, C61

Palabras clave: modelos de predicción, regresión no paramétrica, Técnicas de optimización

## Agradecimientos

El presente trabajo de investigación fue una elaboración gracias a la supervisión del Dr. José Manuel Bravo Caro, a quien me gustaría expresar mi más profundo agradecimiento por su ayuda y colaboración en cada momento, para que esta investigación se haya llevado a cabo.

A mis padres y mis hermanos, por su apoyo incondicional durante todos los años de mi vida de estudiante gracias, con su cariño todo ha sido mucho más fácil.

A los maestros que fueron parte importante dentro del proceso de enseñanza, a la universidad Universidad de Huelva y Universidad Internacional de Andalucía, por ayudar a expandir mis conocimientos y abrirme campo en el apasionante mundo de la analítica de datos.

# Tabla de Contenido

Introducción	p.1
1.- Marco Contextual	p.4
1.1. Energía Fotovoltaica	p.4
1.2. Energía Fotovoltaica en Bélgica	p.5
2.- Conceptos Teóricos	p.7
2.1. Predicción	p.7
2.1.1. Métodos de Regresión	p.8
2.2. Optimización	p.11
2.2.1. Proceso de resolución de un problema de optimización	p.13
2.2.2. Optimización Convexa	p.18
2.2.3. Problema de mínimos cuadrados	p.19
2.3. Restricciones	p.21
2.3.1. Restricciones de igualdad	p.21
2.4. Cálculo del error de predicción: Método de validación cruzada	p.23
2.4.1. Método de validación Leave-One-Out	p.23
3.- Aplicación Práctica	p.26
3.1. Formulación del Problema	p.26
3.2. Predictores usados para comparación	p.28
3.2.1. Modelos de persistencia	p.28

3.2.2. Mínimos Cuadrados Ordinarios	p.29
3.3. Base de Datos	p.29
3.4. Elección del regresor	p.31
3.5. Elección del Gamma óptimo	p.31
3.6. Generación de predicciones	p.35
3.6.1. Predictor de optimización	p.35
3.6.2. Predictor Mínimos Cuadrados Ordinarios	p.37
3.6.3. Predictor persistente	p.38
3.7. Resultados	p.40
4.- Conclusiones	p.42
Referencias	p.43

## Introducción

La analítica predictiva de datos resulta un aspecto fundamental en multitud y muy diferentes ámbitos, ya que permite anticipar escenarios futuros a través de la detección de patrones de comportamiento de los datos. Puesto que el fin de la analítica predictiva es, por lo tanto, dar la mejor aproximación posible sobre el comportamiento futuro de las variables objeto de estudio que miden relaciones del mundo real, resulta evidente su utilidad como herramientas de apoyo a la toma de decisiones a cualquier nivel. Este aspecto toma especial relevancia en el ámbito de la dirección estratégica y de operaciones de empresas cuya actividad productiva dependa en gran medida de estimaciones para llevarse a cabo.

Un caso específico y muy interesante es el de las empresas de producción eléctrica, puesto que como la energía eléctrica no puede ser almacenada, debe de ser producida en una cantidad suficiente para satisfacer la demanda, pero sin incurrir en sobreproducción. En ambos escenarios, una estimación deficiente de la demanda energética resulta en grandes costes económicos, ya que, si la empresa no logra producir la cantidad demandada, estará sujeta a penalizaciones en el mercado eléctrico con la subsecuente pérdida de cuota de mercado y si, por otro lado, la empresa produce más energía de la necesaria, no conseguirá venderla y no se sufragarán los gastos de producción de la misma.

Así pues, en casos como el anteriormente expuesto resulta clara la necesidad real de contar con métodos de predicción adecuados que proporcionen información lo más precisa posible sobre eventos futuros.

En esta investigación se proporciona un nuevo modelo de predicción basado en métodos no paramétricos, en el cual, a diferencia de otros métodos no paramétricos existentes en la literatura, el peso relativo de las observaciones en la estimación se obtiene de la suma ponderada de un problema de optimización convexo, en el cual se incorpora un hiperparámetro de afinación que permite modificar el comportamiento de la función objetivo a optimizar.

Entendiendo entonces que en esta investigación la predicción de valores se hará utilizando un conjunto de datos a través de un enfoque de métodos no paramétricos, resulta necesario explicar la elección de este tipo de métodos en lugar de métodos paramétricos. En el análisis de los métodos paramétricos se establece una hipótesis fija sobre la relación funcional entre las variables exógenas y endógenas y una vez modelizada dicha relación, se extrapola a nuevos datos de entrada para generar predicciones sobre la variable de salida. En los métodos no paramétricos, sin embargo, se asume una relación de dependencia entre estas variables, pero en lugar de intentar obtener una hipótesis fija a priori sobre la estructura del modelo, dicha relación se extrae de la información contenida en los propios datos, con el objetivo de captar el patrón de comportamiento subyacente en los mismos y utilizar este patrón para generar estimaciones de valores futuros para la variable dependiente.

Una de las principales características de los métodos paramétricos es que, una vez modelada la relación funcional de las variables para un problema en concreto, esta puede usarse para la predicción de datos de entrada futuros. Además, los métodos paramétricos permiten explicar e interpretar la relación funcional entre las variables del modelo. Sin embargo, aunque este tipo de métodos son muy útiles para establecer relaciones de causalidad y dependencia entre variables, la rigidez de su estructura implica establecer fuertes hipótesis de relación entre un conjunto de variables que podrían estar explicando solamente parte del modelo y, en consecuencia, estar dando una visión demasiado simplificada de un sistema mucho más complejo.

Por su parte los métodos no paramétricos, al no establecer una hipótesis fija para la relación existente entre las variables de un modelo, no permiten la interpretación de la relación funcional de una forma tan clara como ocurre con los métodos paramétricos. A pesar de esto, puesto que no se establece una forma fija para la relación, se permite una mayor flexibilidad a la hora de intentar captar la esencia de los datos, lo cual resulta en un aumento de la exactitud de las predicciones.

La estructura que sigue esta investigación para alcanzar el objetivo deseado y mostrar al lector de una forma más detallada el contenido de esta, queda dividida en cuatro capítulos. El primer capítulo contempla el marco contextual del problema de investigación propuesto, el segundo capítulo se centra en el marco conceptual teórico de los métodos predictivos, el tercer capítulo

consiste en una aplicación práctica y el cuarto y último capítulo está formado por las conclusiones de esta investigación.

De esta forma, el primer capítulo de este documento pretende dar un planteamiento y una visión general del problema que se pretende abordar. Este capítulo a su vez está subdividido en dos secciones: en la primera parte de este capítulo se aborda el tema de la energía fotovoltaica y una segunda parte donde se hace referencia al mercado de energía en Bélgica

En el segundo capítulo una descripción de los principales conceptos que nos ayudara a comprender el desarrollo práctico de esta investigación.

El tercer capítulo está comprendido por el desarrollo práctico de los conceptos básicos desarrollados en el segundo capítulo, además de presentar a detalle la descripción de las herramientas de predicción empleadas para nuestra investigación.

En el último capítulo desarrollamos las conclusiones de la investigación y los resultados obtenidos en esta de esta investigación.

# Capítulo 1 Marco Contextual

## 1.1 Energía Fotovoltaica

La energía es de vital importancia para las personas para el desarrollo de sus actividades. Muchos países celebran reuniones y con la agenda principal energética, siendo uno de los objetivos buscar un equilibrio entre la producción y demanda de la misma. Para buscar una solución se investigan formas de poder generar y usar la misma de forma más eficiente.

La energía solar fotovoltaica, es una fuente de energía cuya producción es de origen renovable, obtenida directamente a partir de la radiación solar. Este tipo de energía se utiliza generalmente para producir energía a gran escala a través de redes de distribución, aunque también permite alimentar innumerables aplicaciones y aparatos autónomos, abastecer refugios de montaña o viviendas aisladas de la red eléctrica. Debido a que el consumo de energía aumenta constantemente las formas de obtenerlo se fueron haciendo más sofisticadas y baratas, por lo cual obtenerla ahora se puede de forma más eficiente.

Los sistemas de energía de energía fotovoltaica tienen lugar como la fuente más dominante entre las tecnologías de energía renovable. La razón más importante es que es energía ilimitada y limpia de los sistemas de energía solar. Muchos estudios muestran que los sistemas de energía fotovoltaica tendrán una participación importante en la electricidad del futuro [1].

Desde programas de incentivo económico y producción para el autoconsumo han sido impulsados por varios gobiernos, para de esta manera reducir la emisión de gases de efecto invernadero. Es así que este tipo de técnicas cada vez más novedosas en el mercado de energía ha ido surgiendo, es el caso de Bélgica.

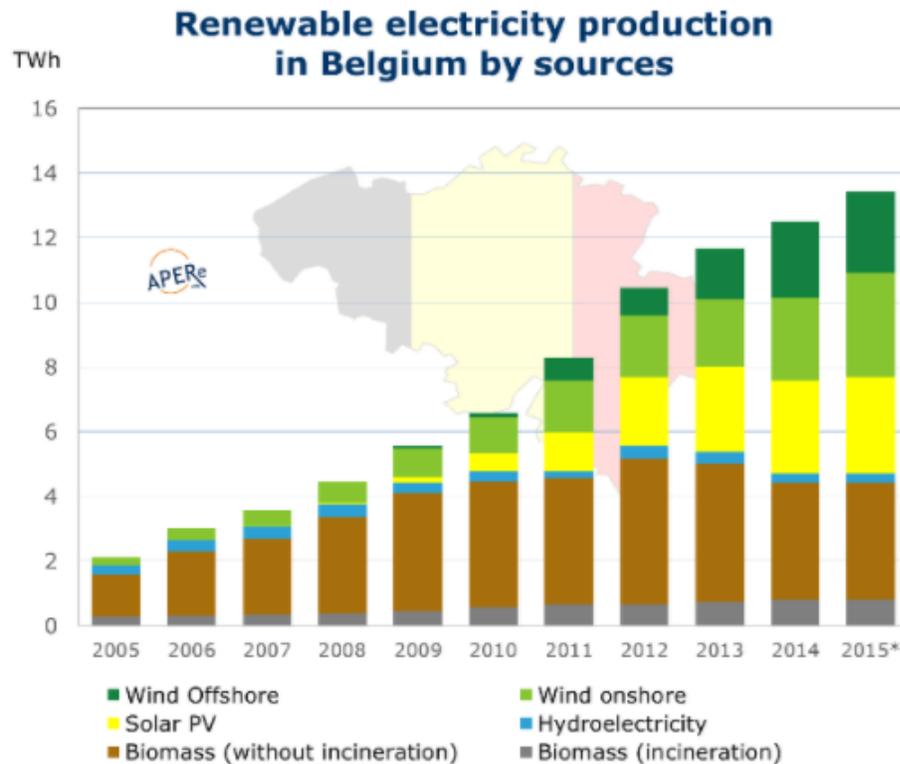
## 1.2 Energía fotovoltaica en Bélgica

Durante los diez últimos años Bélgica ha mostrado un gran crecimiento en la participación de las energías renovables en su consumo final de energía del 2% en el 2005 al 8% en 2014 y 14 % en el 2016.

Las responsabilidades de la política energética en Bélgica se comparten entre el gobierno federal y las tres regiones (Bruselas, Flandes y Valonia). La competencia federal incluye, en términos generales, todos los aspectos relacionados con el suministro de energía y las normas de cada producto. Las regiones gestionan la distribución de energía, incluyendo precios, fuentes de energía renovable, promoción del uso eficiente de energía.

La energía solar en Bélgica alcanzó una capacidad instalada de 3.228 MW de potencia que genera 2.865 GWh de electricidad en 2015. En 2015, la energía solar fotovoltaica representó alrededor del 4% de la demanda eléctrica total de Bélgica, la 4ª cifra de penetración más alta del mundo, aunque el país está un poco por detrás de los líderes Alemania, Italia y Grecia en entre 7% y 8% de la demanda de electricidad, para el 2016 se presentó un escenario similar. Además de incorporar varias y novedosas plantas de generación de energía fotovoltaica con paneles solares en zonas amplias y también la incorporación de las mismas en edificios.

La producción de energía en Bélgica es impulsada por los esquemas regionales de respaldo de certificados verdes iniciados en 2001, la electricidad renovable ha experimentado un aumento real en la producción de capacidad debido al creciente número de turbinas eólicas terrestres y marinas, instalaciones fotovoltaicas solares y plantas de electricidad térmica con biomasa. Bélgica ocupa el cuarto lugar de la UE-28 en 2015 con la mayor capacidad de potencia acumulada de energía eólica offshore. Además, el desarrollo eólico en tierra ha dado lugar a un gran movimiento ciudadano de cooperativas: en 2015, 60 000 ciudadanos habían invertido en acciones de cooperativas de energía renovable (estimación APERe) [1].



Producción de electricidad renovable en Bélgica por fuentes (APERe)

En la grafica se puede observar en color amarillo la evolución de la energía fotovoltaica en Bélgica como se ha ido mencionando va teniendo mayor relevancia y por lo tanto mayor producción desde el 2009.

En esta investigación se va a realizar un análisis de predicción de la producción de energía fotovoltaica para Bélgica, los datos de análisis fueron los datos operativos relacionados con la producción de energía fotovoltaica para el mes de octubre de 2017. Los datos para tal fin se obtuvieron de la página web de Elia en colaboración con FEBEG (Federación Belga de Empresas de Electricidad y Gas), Elia es la mayor empresa de energía eléctrica de Bélgica, fundada en 28 de junio de 2001, y opera de forma regular en este País.

## Capítulo 2 Conceptos Teóricos

### 2.1 Predicción

La predicción es una actividad crucial en diversas áreas, de forma que podemos decir que la predicción es un anuncio sobre algo que pasará a futuro. De hecho, es fácil comprender que el conocimiento de un acontecimiento futuro permite prepararse de forma adecuada para hacerle frente. Este es un aspecto importante para las empresas ya que les permite planificar mejor su producción y cubrir una determinada demanda sin quedarse corto ni con exceso, como anteriormente se venía indicando.

Para llevar a cabo la previsión de acontecimientos futuros, se utilizan técnicas de análisis de datos que ayudan a interpretar conjuntos de datos y que permiten extraer información representativa sobre las relaciones subyacentes entre los datos. Así, es posible obtener predicciones en momentos no observados en el futuro o pasado.

Los nuevos tiempos están marcados por la era digital, la globalización, la ingente cantidad de datos cada vez más sofisticados y de diversas áreas generados con una frecuencia cada vez mayor. Ante este nuevo escenario surge una mayor necesidad de analizar estos datos con el fin de extraer información valiosa de ellos y generar conocimiento a partir de estos. Como consecuencia de ello, el campo de la analítica de datos ha sufrido un desarrollo muy importante en los últimos años, recibiendo aportes de muy distintas disciplinas como son las ciencias de la computación, las matemáticas o la economía entre otras, lo cual ha dado lugar a la aparición de nuevas y sofisticadas herramientas para el tratamiento de datos. Dentro de este complejo contexto surge el denominado Machine Learning o Aprendizaje Automático, el cual consiste básicamente en el uso de algoritmos para la identificación automática de patrones en los datos que posteriormente puedan ser utilizados para estimar futuros comportamientos de las variables objeto de estudio.

Los algoritmos de Machine Learning a su vez se clasifican en algoritmos de aprendizaje supervisado y de aprendizaje no supervisado. El aprendizaje supervisado, es una técnica para deducir una función a partir de un conjunto de datos de entrenamiento, los cuales a su vez están compuestos por dos componentes, uno de entrada y otro de salida. Por otro lado, el aprendizaje no

supervisado es un paradigma completamente diferente, en el cual solamente se tienen datos de entrada, pero no de salida. El objetivo del aprendizaje no supervisado es la detección de patrones y similitudes entre los datos de entrada.

Así pues, en lo que a métodos de predicción se refiere, los algoritmos que resultarán útiles en este campo serán los algoritmos de aprendizaje supervisado, los cuales a su vez se subdividen en métodos de regresión y de clasificación. La diferencia entre los métodos de regresión y de clasificación es que, en los primeros, la variable dependiente a predecir es una variable de tipo continuo, por lo que los valores de salida de nuestras predicciones serán siempre valores continuos. En el caso de la clasificación, sin embargo, la variable dependiente es de tipo discreto, ya que intenta representar estados mutuamente excluyentes entre sí o grupos de clasificación, por lo que los valores que obtendremos de aplicar un método de clasificación serán siempre valores discretos que representarán la pertenencia a un grupo de clasificación concreto.

Siendo el propósito de estudio de los modelos de regresión construir modelos matemáticos que permitan explicar la relación de dependencia existente entre una variable dependiente y una o más variables independientes, podemos utilizar estos modelos como herramienta para predecir nuevos valores de la variable dependiente a partir de valores concretos que tomen las variables independientes. Con el objetivo de realizar predicciones, para esta investigación haremos uso del aprendizaje supervisado y dentro de esta estudiaremos el método de la regresión.

### 2.1.1 Métodos de Regresión

Uno de los aspectos más relevantes dentro del campo de la predicción ha venido siendo el análisis de los modelos de regresión, ya que nos permiten cuantificar la forma en la que influyen una o varias variables explicativas  $X$  en una variable respuesta  $Y$ . De forma general dicha relación puede expresarse como:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \dots, n$$

En la ecuación anterior,  $y_i$  es la variable de salida,  $x_i$  hace referencia al vector de variables de entrada,  $\varepsilon_i$  es la perturbación aleatoria del modelo y  $f(x_i)$  es la relación entre las variables de entrada y la variable de salida, es decir, la función de regresión.

Existen dos vertientes dentro del campo de los métodos de regresión: los métodos paramétricos y los no paramétricos. En ambos casos el fin último es realizar predicciones cuantificadas de forma continua para cierta variable de estudio, pero la forma de llegar a esa predicción es muy diferente entre cada uno de los métodos.

Por un lado, los métodos basados en regresión paramétrica asumen que la función de regresión tiene una forma predeterminada y el objetivo de estos métodos es obtener una función adecuada que modele la relación entre las variables del modelo. Por otro lado, los métodos basados en regresión no paramétrica no asumen ninguna forma predeterminada para la función de regresión, sino que solamente consideran una hipótesis de suavidad sobre la función de regresión.

Aunque los métodos paramétricos resultan muy útiles para explicar las relaciones que existen entre las variables de un sistema en concreto, su principal desventaja radica en que la función subyacente real desconocida puede tener una forma muy compleja que resultaría muy difícil de identificar a través de las herramientas que estos métodos nos aportan. Además, en algunos casos resulta demasiado arriesgado dar una suposición firme sobre la forma funcional de los datos en cuestión, ya que estos métodos quedan restringidos a la forma especificada, por lo que cuando se presentan problemas muy complejos que necesitan de una exactitud de las estimaciones muy alta, estos métodos pueden presentar un ajuste deficiente en la práctica.

Por su parte, los métodos no paramétricos buscan ajustarse mejor a los datos de entrenamiento en la construcción de la función, a la vez que se mantiene la capacidad para generalizar a datos no conocidos con anterioridad. Como tales, son capaces de adaptarse a una gran cantidad de formas funcionales, ya que no imponen ningún tipo de restricción en cuanto a la forma funcional de los datos, cosa que como ya se ha comentado, sí que ocurre con los métodos paramétricos.

La diferencia de los métodos paramétricos y los no paramétricos estriba, por lo tanto, como hemos visto, en la forma de tratar la relación entre las variables del modelo. Así, lo que cambiará al usar

un tipo de métodos u otros será precisamente el cálculo de la función de regresión, puesto que en el caso paramétrico esta será construida como resultado de la obtención de los parámetros que minimizan el error para una familia de funciones elegida, mientras que en el caso no paramétrico esta función que relaciona las entradas y salidas del modelo se generará a partir de un estimador.

El uso de un tipo u otro de métodos de regresión estará relacionado con el objetivo que se persiga: mientras que los métodos paramétricos son muy útiles para establecer hipótesis sobre la relación entre variables, los métodos no paramétricos son más precisos a la hora de generar estimaciones, aunque en detrimento de la interpretabilidad. Además, la regresión no paramétrica es imprescindible cuando la variable que se pretende predecir es imposible o muy costosa de medir.

Siendo nuestro objetivo el de obtener un método predictivo que nos proporcione predicciones con la mayor exactitud posible, haremos uso por lo tanto de técnicas que se encuadran dentro del terreno de los métodos no paramétricos.

Estos métodos son computacionalmente costosos de implementar, por lo que se hará uso de herramientas computacionales que nos ayuden a obtener los resultados, a través de la programación de algoritmos de predicción no paramétricos, para posteriormente poder comparar los resultados con aquellos obtenidos con otros métodos de predicción.

Los métodos no paramétricos que se estimarán en este apartado consideran que la predicción,  $\hat{y}_k$ , se construye como una media ponderada a las  $\hat{y}_i$  conocidas, obtenidas a través de un proceso de asimilación que corresponde a un problema de optimización. Este método no hace suposiciones sobre la distribución que siguen los datos, a diferencia, de un método basado en técnicas de regresión paramétricas, por lo que el mejor modelo de los datos son los propios datos.

Dentro de los estimadores no paramétricos más sofisticados necesitamos siempre obtener un vector de pesos que establezca el nivel de ponderación de cada uno de los datos disponibles en función de la similitud de este con el dato a predecir. Este vector de pesos se puede obtener de muy diferentes formas, siendo una de las más empleadas el uso del estimador de Nadaraya-Watson [2, 3], que básicamente usa un método Kernel de estimación de funciones de densidad. La novedad

que se propone en esta investigación en concreto, el vector de pesos se obtiene de la resolución de un problema de optimización convexa con restricciones de igualdad.

Al ser la resolución de un problema de optimización con restricciones la base sobre la que se sustenta la esencia de este modelo predictivo propuesto, resulta conveniente exponer los conceptos básicos encuadrados en el marco de los problemas de optimización.

## 2.2 Optimización

Dentro del campo de la informática y las matemáticas, la optimización se define como la selección del mejor elemento a partir de un conjunto de alternativas disponibles. En el caso más simple, un problema de optimización consiste en maximizar o minimizar una función real eligiendo sistemáticamente los valores de entrada desde dentro de un conjunto permitido y calculando el valor de la función. De manera más genérica podríamos decir que un proceso de optimización se fundamenta en encontrar los mejores valores posibles de alguna función objetivo atendiendo a las posibles restricciones del modelo que se puedan plantear, lo cual incluye una variedad de diferentes tipos de funciones objetivo y diferentes tipos de dominios.

La capacidad de adaptación a no linealidades y discontinuidades cada vez más eficiente hace de estos métodos sean usados en las últimas décadas, siendo aplicadas con éxito en muchas áreas. [4].

Así pues, un problema de optimización tendría la siguiente forma:

$$\begin{aligned} & \text{minimizar } f_0(x) \\ & \text{sujeto a } f_i(x) \leq b_i, \quad i = 1, \dots, m. \quad (1.1) \end{aligned}$$

Aquí el vector  $x = (x_1, \dots, x_n)$  es la variable de optimización de problema, la función  $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$  es la función objetivo, las funciones  $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m$ , son las funciones de restricción (de desigualdad en este caso), y las constantes  $b_1, \dots, b_m$  son los límites para las restricciones. Un vector  $x^*$  se llama óptimo, o una solución del problema (1.1), si tiene el menor valor objetivo de entre todos los vectores que satisfacen las restricciones

Generalmente consideramos familias o clases de problemas de optimización que se caracterizan por formas particulares de la función objetivo y las funciones de restricción. Como un ejemplo importante el problema de optimización (1.1) se llama programa lineal si el objetivo y funciones de restricción  $f_0, \dots, f_m$  son lineales, es decir, satisfacen:

$$f_i(\alpha x + \beta y) = \alpha f_i(x) + \beta f_i(y) \quad (1.2)$$

Para todo,  $y \in \mathbb{R}^n$  y todo  $\alpha, \beta \in \mathbb{R}$ . Si el problema de optimización no es lineal, es llamado programación no lineal.

Por lo tanto, el problema de optimización de la ecuación (1.1), es una forma de elección del mejor valor que pertenece a  $\mathbb{R}^n$  de un conjunto de elecciones de valores candidatos. La variable  $x$  representa la elección hecha, las restricciones  $f_i(x) \leq b_i$  representan requisitos firmes o especificaciones que limitan las opciones posibles, y el valor de la función objetivo representa el costo de elegir  $x$ . Entonces podríamos decir que una solución del problema de optimización (1.1) corresponde a una elección que tiene el coste mínimo (o utilidad máxima) entre todas las opciones que satisfagan todas y cada una de las condiciones impuestas por el modelo.

En el ajuste de datos, la tarea es encontrar un modelo, de una familia de modelos potenciales, que mejor se adapte a algunos datos observados e información previa. Aquí las variables son parámetros en el modelo, y las restricciones pueden representar información previa o límites requeridos en los parámetros (como la no negatividad). La función objetivo podría ser una medida de inadaptación o error de predicción entre los datos observados y los valores predichos por el modelo, o una medida estadística de la improbabilidad o la inverosimilitud de los valores de los parámetros.

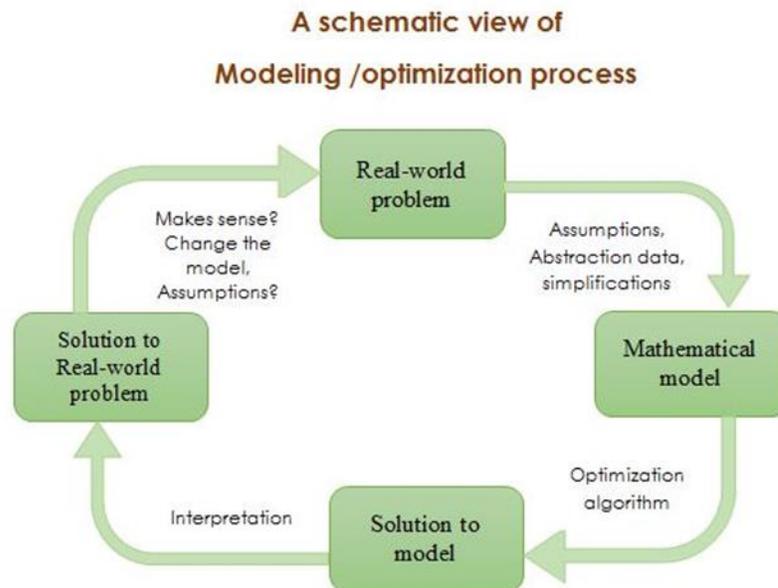
### 2.2.1 Proceso de resolución de un problema de optimización

El modelado de un problema de optimización, consiste en varias etapas diferenciadas. Estas consisten básicamente en traducir al lenguaje matemático una problemática concreta que modele un problema del mundo real, para a continuación pasar a aplicar un algoritmo adecuado para su

resolución. expresar un problema práctico en forma de problema de optimización. El siguiente paso en el proceso consiste en la realización de un análisis de la eficiencia de estos algoritmos de resolución, con el objetivo de estudiar la bondad de los algoritmos de optimización y ver cómo de rápido convergen a la solución óptima.

El problema de optimización se debe ajustar a ciertas funciones a través de la predicción o parámetros a través del modelado. Como ya se viene comentando, el objetivo es hallar una solución óptima desde el punto de vista del modelo, ya que cuanto más se adapte el modelo desarrollado al mundo real, más cerca estarán las soluciones óptimas del modelo y las soluciones óptimas reales. Así podríamos decir que, dado un conjunto de ejemplos, el problema consistirá en obtener un único punto que represente al grupo.

De esta forma, un problema de optimización debe modelarse como se muestra a continuación, en el siguiente esquema:



### Modelización:

La forma básica de un problema de optimización sería algo similar a la siguiente expresión:

$$J^* \min_x J(x)$$

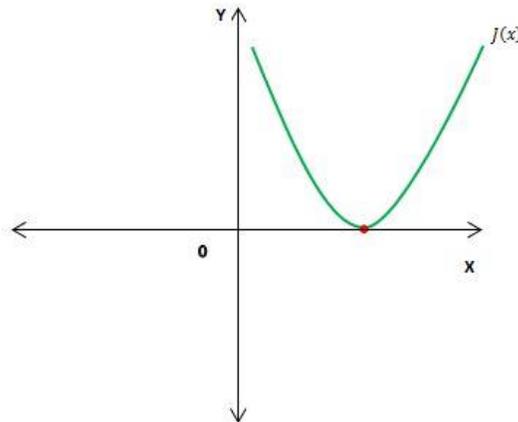
Donde:

- $x$  es el representante buscado la variable de decisión del problema de optimización.
- $J(x)$  es la función objetivo del problema a resolver
- $x_i$  son los datos disponibles para el problema en cuestión

Para poder obtener la solución al problema de optimización realizaríamos los siguientes pasos:

1. Simplificar el problema suponiendo un único dato. En este caso, el representante optimo debería de ser el dato conocido.
2. Diseñar una función objetivo como una función de coste

El valor mínimo que debe de tomar es 0 si la solución representa al conjunto de datos de forma perfecta. Donde  $J^* = 0$ . A continuación se muestra una representación gráfica:

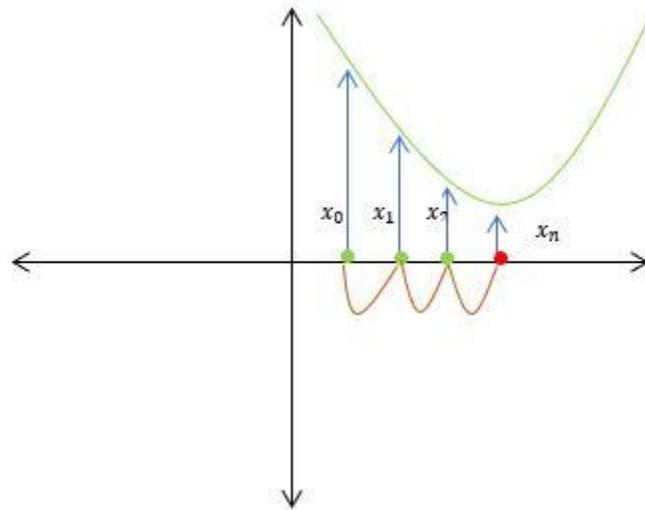


El problema se complica sin embargo cuando tenemos varios datos donde a la vez se cuenta con varias funciones de costes que hay que combinar. Una posibilidad entonces sería sumar todas las funciones de coste, con el objetivo de obtener un óptimo global.

Entonces para obtener la solución óptima podemos hacerlo de dos formas: a través de manipulación algebraica (expresión explícita) o manipulación numérica (método iterativo). Este último es ampliamente usado en los algoritmos implementados sobre herramientas de software y

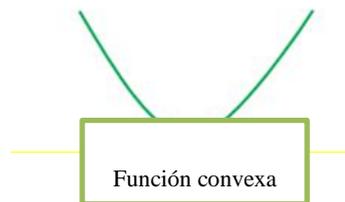
básicamente consiste en: empezar en un punto inicial  $x_0$  y generar de forma iterativa una secuencia de  $x_0, x_1, \dots, x_k$  puntos. Esta secuencia debe de converger a  $x^*$ .

Gráficamente se tiene:



Tanto el valor de la solución óptima y la convergencia de los métodos iterativos al óptimo global dependen de la convexidad de la función objetivo. Tal como habíamos mencionado, nos encontramos ante un problema de no linealidad por la forma de la figura anterior y por lo expuesto anteriormente podemos decir que nos encontramos con un problema de programación no lineal. Este problema puede representarse como se detalla a continuación, de acuerdo a las características que presenten la función objetivo y las restricciones.

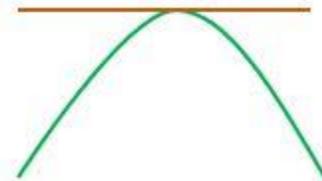
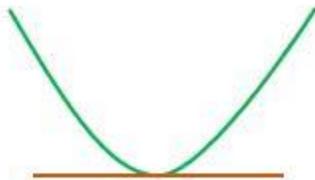
$$\begin{cases} \min f(\vec{x}) \\ \vec{x} \in D \end{cases} \text{ cuando la gráfica de } f \text{ es}$$



$$\begin{cases} \max f(\vec{x}) \\ \vec{x} \in D \end{cases}$$
 cuando la gráfica de  $f$  es



En ambos casos, el problema de optimización se reduce a calcular el punto en el que la derivada es nula, ya que esta situación representa un punto de inflexión, lo cual siempre estará asociado a un máximo o mínimo en función de si la función objetivo es convexa o cóncava:



En el primer caso se obtiene un mínimo global o absoluto, y en el segundo caso un máximo global o absoluto.

Un problema del tipo

$$\begin{cases} \min f(\vec{x}) \\ \vec{x} \in D \end{cases}$$
 donde  $f$  es una función convexa y  $D$  un dominio convexo.

Se llama problema convexo

Paralelamente

$$\begin{cases} \max f(\vec{x}) \\ \vec{x} \in D \end{cases} \quad \text{donde } f \text{ es una función cóncava y } D \text{ un dominio convexo.}$$

Se llama problema cóncavo o problema convexo para maximizar.

La elección del método de resolución y del algoritmo adecuado para ello se realizará en función de si nuestra función objetivo corresponda a una función lineal o no. Así, podremos distinguir dos tipos de problemas:

Si la función objetivo  $f$  es lineal y el espacio restringido es un politopo, el problema es de programación lineal y puede resolverse utilizando algoritmos adecuados para este tipo de problemas. En cambio, si la función objetivo es cóncava (problema de maximización), o convexa (problema de minimización) y el conjunto de restricciones es convexo, entonces se puede utilizar el método general de optimización convexa.

En esta investigación en concreto, por las particularidades intrínsecas de la función de distribución de los pesos para las observaciones, el problema de optimización a resolver para obtener el vector de pesos óptimo resulta ser un problema de tipo convexo. Por ello, estudiaremos con detalle la optimización convexa, con el objetivo de dar al lector una visión más amplia sobre la problemática que encierra el problema de optimización en cuestión, el cual, como ya se comentó anteriormente, es el elemento diferenciador con respecto a otros métodos de obtención del peso relativo de las observaciones que ya existen en la literatura y son ampliamente usados continuamente.

### 2.2.2 Optimización Convexa

La optimización convexa incluye problemas como los mínimos cuadrados y problemas de programación lineal. Es bien sabido que los mínimos cuadrados y los problemas de programación lineal surgen muy frecuentemente en una gran variedad de aplicaciones, y se pueden resolver de una forma muy eficiente numéricamente. Hay grandes ventajas para reconocer o formular un

problema como un problema de optimización convexo, siendo la ventaja primordial la eficiencia y fiabilidad de los métodos de resolución para este tipo de problemas, usando métodos como el método de punto interior, por ejemplo.

Los conjuntos y funciones convexas tienen algunas propiedades que los hacen especiales como problemas de optimización, como son:

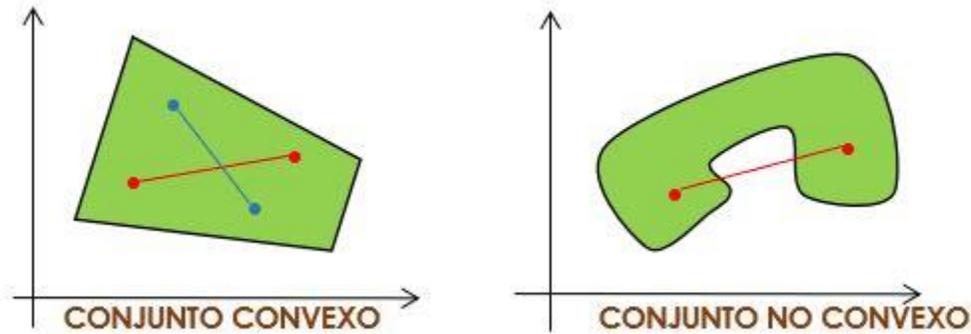
- Una función convexa no tiene mínimos locales que no sean globales.
- Un conjunto convexo tiene un interior relativo no vacío.
- Un conjunto convexo está conectado y tiene direcciones factibles en cualquier punto.
- Una función convexa es continua dentro del interior de su dominio, y tiene buenas propiedades de diferenciación.

Entonces podríamos replantear el problema de optimización convexa como aquel en el que la función objetivo y las funciones de restricción son convexas, lo que significa que satisfacen la siguiente desigualdad:

$$f_i(\alpha x + \beta y) \leq \alpha f_i(x) + \beta f_i(y) \quad (1.3)$$

Para todo  $x, y \in \mathbb{R}^n$  y todo  $\alpha, \beta \in \mathbb{R}$  con  $\alpha + \beta = 1, \alpha \geq 0, \beta \geq 0$ . Comparando (1.2) y (1.3), vemos que la convexidad es más general que la linealidad: la desigualdad reemplaza la igualdad más restrictiva, y la desigualdad debe mantenerse solo para algunos valores de  $\alpha$  y  $\beta$ . Como cualquier programa lineal es, por lo tanto, un problema de optimización convexa, podemos considerar la optimización convexa como una generalización de la programación lineal.

La característica que otorga a un conjunto de convexidad el que dicho conjunto contiene al segmento que pasa por dos puntos diferentes que pertenecen a este. Para ilustrar mejor lo mencionado haremos una representación gráfica como se muestra en la siguiente figura.



Algunas expresiones convexas pueden ser representaciones matemáticas de algunas figuras como: bolas, elipsoides, hiperplanos, semiespacios, conos o polítopos.

### 2.2.3 Problema de Mínimos cuadrados

Como ya se ha comentado anteriormente la optimización convexa incluye problemas como los mínimos cuadrados, es decir es un problema de optimización sin restricciones, (si  $m = 0$ ) y un objetivo que es la suma de cuadrados de la siguiente forma  $a_i^T x - b_i$ :

$$\text{minimizar } f_0(x) = \|Ax - b\|_2^2 = \sum_{i=1}^k (a_i^T x - b_i)^2 \quad (1.4)$$

Aquí  $A \in \mathbb{R}^{k \times n}$  (con  $k \geq n$ ),  $a_i^T$  son las filas de  $A$ , y el vector  $x \in \mathbb{R}^n$  es la variable de optimización.

Para la resolución de los problemas de mínimos cuadrados concretamente la ecuación (1.4), se podría reducir a la solución de un conjunto de sistema de ecuaciones lineales como:

$$(A^T A)x = A^T b,$$

Entonces se tiene la siguiente solución analítica  $x = (A^T A)^{-1} A^T b$ . Para este tipo de problemas se han ido desarrollando e implementando diversos algoritmos y resolverlo con una precisión cada vez más alta.

El problema de los mínimos cuadrados es la base del análisis de regresión, control óptimo y muchos métodos de estimación de parámetros y ajuste de datos. Reconocer un problema de optimización como un problema de mínimos cuadrados es sencillo; solo necesitamos verificar que

el objetivo sea una función cuadrática. Mientras que el problema básico de mínimos cuadrados tiene una forma fija simple, varias técnicas estándar se utilizan para aumentar su flexibilidad en las aplicaciones.

En mínimos cuadrados ponderados, el costo ponderado de mínimos cuadrados

$$\sum_{i=1}^k w_i (a_i^T x - b_i)^2,$$

donde  $w_1 \dots w_n$  es positiva, se minimiza.

Otra técnica en mínimos cuadrados es la regularización, en la cual los términos adicionales son agregados a la función de coste. En el caso más simple, un múltiplo positivo de la suma de cuadrados de las variables se agrega a la función de coste como se muestra a continuación:

$$\sum_{i=1}^k (a_i^T x - b_i)^2 + \rho \sum_{i=1}^n x_i^2,$$

donde  $\rho > 0$ . Los términos adicionales penalizan valores grandes de  $x$ , y dan como resultado una solución en casos cuando minimizar la primera suma solo no lo hace. El parámetro  $\rho$  es elegido por el usuario para dar la compensación correcta entre hacer la función objetivo general.  $\sum_{i=1}^k (a_i^T x - b_i)^2$  pequeña, mientras se mantiene  $\rho \sum_{i=1}^n x_i^2$  no demasiado grande.

## 2.3 Restricciones

En la mayor parte de los problemas de toma de decisiones están presentes ligaduras entre las variables o limitaciones en las mismas, unas debidas a las ecuaciones del modelo, otras al rango permisible de unas variables y otras debidas a reglas de operación, existencias, etc.

Por ello un problema de optimización a menudo se formula como la forma matemática de la ecuación (1.1), pero para expresarlo con restricciones de igualdad añadiremos un componente de igualdad en la restricción como:

$$\min_x J(x)$$

$$h(x) = 0$$

$$g(x) \leq 0$$

La presencia de restricciones limita el espacio de búsqueda, pero al mismo tiempo, dificulta el encontrar la solución óptima porque se pierden algunos de los criterios de optimalidad como que el gradiente es nulo en el óptimo

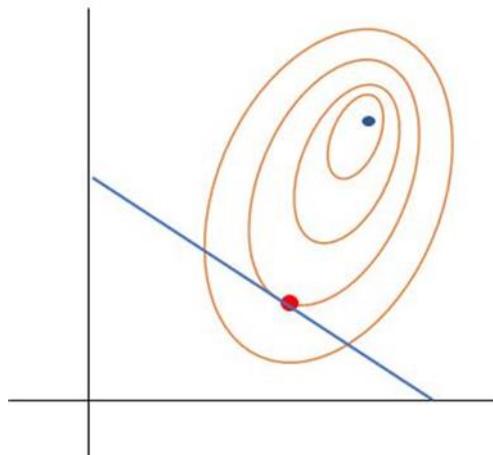
### 2.3.1 Restricciones de igualdad

Hay una categoría de problemas en los cuales las variables de decisión están sujetas solo a un conjunto de ecuaciones de igualdad:

$$\min_x J(x)$$

$$h(x) = 0$$

Así gráficamente podríamos expresarlo de la siguiente forma:



La region factible esta comprendida por los puntos de interseccion entre los puntos de las curvas de nivel y la recta que corresponde a la restriccion de igualdad, pero la solución óptima siempre coincidirá con la intersección entre una curva de nivel y una recta que la corte de manera tangencial.

En este caso para este tipo de problemas de optimización con restricciones de igualdad, el método de los multiplicadores de Lagrange proporciona condiciones necesarias que deben cumplirse en el óptimo. La idea es convertir el problema en otro sin restricciones ampliado en  $m$  variables  $\lambda_j$  (los multiplicadores de Lagrange) tal que su solución coincida en las variables  $x$  con el primitivo y cumpla las restricciones  $h(x) = 0$ .

$$\left. \begin{array}{l} \min_x J(x) \\ h(x)=0 \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \min_{x,\lambda} L(x, \lambda) = \min_{x,\lambda} J(x) + \lambda' h(x) \\ \text{Para todos los } x \text{ que cumplan las restricciones } h(x) = 0 \text{ se verifica:} \\ \min_{x,\lambda} L(x, \lambda) = \min_x J(x) \end{array} \right.$$

Si  $x^*$  es óptimo para el problema original, minimiza  $J(x^*)$  y cumple  $h(x^*) = 0$ , luego también tiene que ser una solución del problema de Lagrangiana  $L$ .

$$\min_{x,\lambda} L(x, \lambda) = \min_{x,\lambda} J(x) + \lambda' h(x) \quad L(x, \lambda) \text{ Lagrangiana}$$

Por lo que la solución del problema ampliado sin restricciones es:

$$\left. \begin{array}{l} \frac{\partial L(x, \lambda)}{\partial x} \Big|_{x^*, \lambda^*} = 0, \quad \frac{\partial L(x, \lambda)}{\partial \lambda} \Big|_{x^*, \lambda^*} = 0 \rightarrow h(x^*) = 0 \end{array} \right\}$$

Con lo cual obtenemos una solución explícita.

## 2.4 Cálculo del error de predicción: Método de validación cruzada

En la presente investigación se hace uso del método de validación cruzada para calcular el error asociado a la predicción.

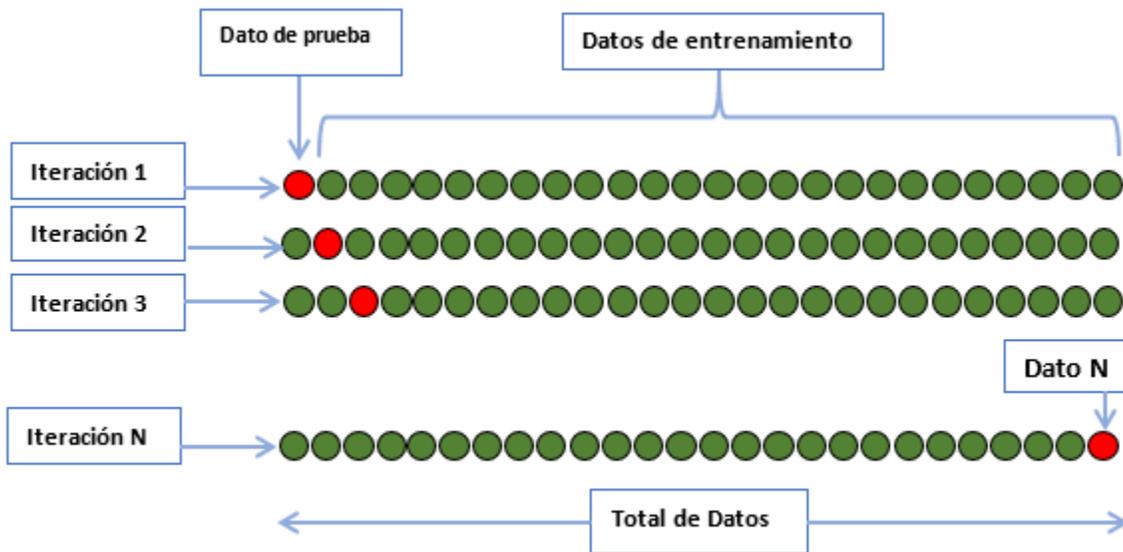
El método de validación cruzada es una técnica utilizada para evaluar los resultados de un análisis estadístico, y garantizar que son independientes entre datos de prueba y de entrenamiento. Consiste en repetir y aplicar la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones de nuestra muestra de datos.

Se utiliza en entornos donde el objetivo principal es la predicción y se quiere estimar la precisión de un modelo que se llevará a cabo a la práctica.

Dentro de los tipos de validación cruzada para el caso de esta investigación en concreto haremos uso del método de validación cruzada con la técnica Leave-One-Out.

### 2.4.1 Método de Validación Leave-One-Out

El método de validación cruzada dejando uno fuera o 'Leave-one-out cross-validation', implica separar los datos de forma que para cada iteración tengamos una sola muestra para los datos de prueba y todo el resto conformando los datos de entrenamiento. La evaluación viene dada por el error, y en este tipo de validación cruzada el error es muy bajo, pero en cambio, a nivel computacional es muy costoso, puesto que se tienen que realizar un elevado número de iteraciones, tantas como N muestras tengamos y para cada una analizar los datos tanto de entrenamiento como de prueba.



La evaluación de la validación cruzada normalmente viene dada por el error obtenido en cada iteración, esta pueda variar en función al número de iteraciones del método de validación y está en función del número de datos con el que se trabaja.

Aunque este método es computacionalmente muy costoso, tal como anteriormente se ha indicado, destaca por la precisión derivada de la exhaustividad en la comprobación del error asociado a las predicciones.

### Cálculo de error en Leave-One-Out

En la validación cruzada dejando uno fuera se realizan tantas iteraciones como muestras (N) tenga el conjunto de datos, de forma que para cada una de las N iteraciones se realiza un cálculo de error. El resultado final lo obtenemos realizando la media aritmética de los N valores de errores obtenidos, según la fórmula:

$$E = \frac{1}{K} \sum_{i=1}^K E_i$$

Donde se realiza el sumatorio de los N valores de error y se divide entre el valor de N.



## Capítulo 3 Aplicación Práctica

En este capítulo se aporta una formulación del problema en cuestión y se muestra el proceso completo para la obtención de resultados, tanto para el predictor que se implementa en la presente investigación, como los demás predictores que servirán como base de comparación de la bondad de ajuste de nuestro predictor con respecto a otros métodos de predicción.

De esta forma, comenzamos en primer lugar aportando una formulación del problema a tratar con el predictor basado en métodos de optimización y posteriormente se da una especificación de los demás predictores que se usan en el proceso de comparación.

### 3.1 Formulación del problema

Partiendo del método de regresión, como ya se indicó, consideremos que una variable dependiente está construida a través de una función de regresión, tal como se muestra en la siguiente ecuación:

$$y_k = f_0(x_k) + e_k \quad (1)$$

Donde  $f_0(\cdot, \cdot)$  es desconocida,  $k$  es el instante de tiempo discreto,  $y_k \in \mathbb{R}$  es la salida del sistema,  $e_k$  representa el valor de componente aleatorio. Por otra parte,  $x_k \in X \subseteq \mathbb{R}^{n_x}$  representa el regresor, el cual puede estar compuesto por entradas o salidas del sistema, o ambas.

De esta forma, el regresor es lo suficientemente genérico para incorporar tanto elementos exógenos como elementos endógenos. No obstante, en la presente investigación, se considerará una serie temporal, por lo que los componentes del regresor serán únicamente combinaciones de salidas del sistema en instantes del tiempo anteriores.

Consideremos ahora el siguiente problema de optimización:

$$\lambda^* = \arg \min \lambda^T \lambda$$

$$s. t. \quad v^T \lambda = 1 \quad (2)$$

Donde  $\lambda \in \mathbb{R}^N$  y  $v \in \mathbb{R}^N$  es un vector de unos. La solución explícita a este problema de optimización es  $\lambda^* = \frac{1}{N}v$ . La media de la muestra podría calcularse por lo tanto como  $\bar{y} = (\lambda^*)^T Y$  con  $Y = [y_1, \dots, y_N]^T$ . En resumen, el problema de optimización (2) proporciona un vector de pesos óptimos que pueden usarse para obtener una estimación de la media.

El enfoque que se da en esta investigación es ampliar estas ideas a los problemas de predicción. Por lo cual una forma adecuada para obtener el predictor es suponer que el sistema desconocido (1) pertenece a una familia parametrizada de sistemas afines, es decir,  $y_k = f(x_k) + e_k = r_k^T \theta + e_k = [x_k^T \ 1] \theta + e_k$  donde  $\theta$  es un vector de parámetros desconocidos y  $e_k$  es un error normal con media cero y varianza  $\sigma^2$ . A través del uso de un conjunto de datos  $D$  y el método de mínimos cuadrados ordinarios (MCO), se puede obtener un vector de parámetros óptimo,  $\theta^*$ .

Si denotamos las matrices  $A_x^T = [x_1, x_2, \dots, x_N]$  y  $A = [A_x v]$ , entonces la expresión conocida de la solución óptima es  $\theta^* = (A^T A)^{-1} A^T Y$ . De esta manera la predicción  $\hat{y}_k = [x_k^T \ 1] \theta^*$  puede ser obtenida.

La idea anteriormente expuesta para el caso de la regresión paramétrica puede extenderse al campo de la regresión no paramétrica a través del siguiente problema de optimización:

$$\begin{aligned} \lambda^* &= \arg \min \lambda^T \lambda \\ \text{s. t. } &A^T \lambda = r_k \end{aligned} \quad (3)$$

De esta forma, el problema de optimización (3) tiene una solución explícita. De hecho teniendo en cuenta la función de Lagrange  $\mathcal{L}(\lambda, \mu) = \lambda^T \lambda + \mu^T (A^T \lambda - r_k)$ , si hallamos  $\frac{\partial \mathcal{L}}{\partial \lambda}(\lambda, \mu) = 0$  obtenemos el vector óptimo  $\lambda^*$  que será  $\lambda^* = -\frac{1}{2} A \mu$ .

Puesto que  $A^T \lambda = r_k$  obtenemos que  $\mu^* = -2(A^T A)^{-1} r_k$  y finalmente tenemos que  $\lambda^* = A(A^T A)^{-1} r_k$ .

Finalmente, la predicción,  $\hat{y}_k$ , puede ser escrita de la siguiente forma  $\hat{y}_k = (\lambda^*)^T Y$ .

En esta investigación en concreto se extiende el anterior problema de optimización al siguiente sistema:

$$\lambda^* = \arg \min \gamma \|w\lambda\|_1 + \lambda^T w\lambda$$

$$s. a. A^T \lambda = r_k$$

Donde la matriz  $W$  introduce localidad en la predicción, lo cual hace que aumente la exactitud de las predicciones.

## 3.2 Predictores usados para comparación

Como anteriormente se viene indicando, es necesario contar con otros métodos predictivos que nos sirvan de base para comparar las predicciones obtenidas con nuestro predictor, puesto que una predicción, si no es comparada con ninguna otra, no es posible cuantificar de una forma completa la bondad del ajuste de las predicciones.

Por ello, para la presente investigación se proponen dos métodos de predicción diferentes, el primero de ellos consistente en el método de Mínimos Cuadrados Ordinarios y el otro en un predictor persistente.

### 3.2.1 Modelos de persistencia

Este tipo de modelos se basan en considerar que un dato concreto del pasado es similar al dato que se intenta predecir, por las características intrínsecas de los propios datos a predecir, por lo que la predicción es el valor del propio dato del pasado. Por lo tanto, en este método las predicciones se generan mediante un mecanismo automático establecido a priori, es decir, es un procedimiento de predicción que repite de forma mecánica un comportamiento pasado.

Se trata de la forma más simple de modelos predictivos, ya que solamente requieren del conocimiento de un dato concreto del pasado para generar las predicciones. De forma genérica, siendo  $k$  un instante futuro del tiempo, sobre el cual se quiere realizar una predicción, tendremos:

$$\hat{y}_k = \hat{y}_{k-t}$$

Los pronósticos mediante este modelo proporcionan un punto de referencia con el que se pueden comparar modelos más sofisticados. Es uno de los algoritmos más utilizado debido a su sencillez y fácil implementación a la hora de hacer predicciones.

La desventaja que presenta este tipo de modelos, es que su capacidad es muy limitada por lo que no suelen utilizarse más que como referencia para evaluar la calidad de métodos más complejos.

### 3.2.2 Mínimos Cuadrados Ordinarios

En secciones anteriores ya se desarrolló el método de los Mínimos Cuadrados Ordinarios, pero recordemos que, para un vector de regresión  $X$  y un vector de términos pasados de la variable endógena,  $Y$ , la obtención de los parámetros óptimos se realizaba como:

$$\theta^* = (A^T A)^{-1} A^T Y$$

De esta forma, las predicciones se obtendrán como:

$$\hat{y}_k = [x_k^T \ 1] \theta^*$$

## 3.3 Base de Datos

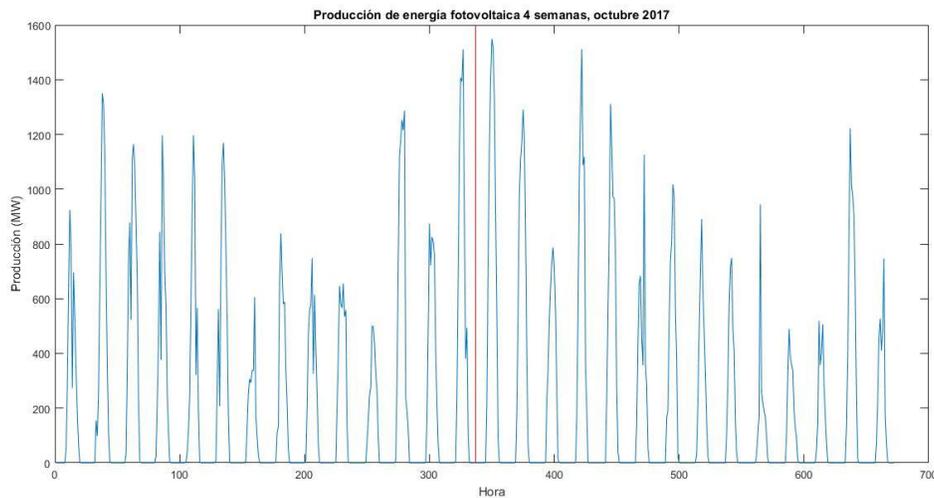
Como ya se ha comentado anteriormente, para llevar a cabo la aplicación práctica, se ha hecho uso de una base de datos que contiene una serie temporal de producción de energía fotovoltaica para el mes de Octubre de 2017. De esta base de datos se ha hecho uso de las primeras 4 semanas, de las cuales las 2 primeras son utilizadas como conjunto de entrenamiento para la obtención del hiperparámetro ( $\gamma$ ) óptimo.

Esta base de datos se ha obtenido a través de la página web de Elia, la mayor compañía del sistema eléctrico belga.

Para la generación de predicciones han sido necesarias tareas previas de limpieza de dicha base de datos, con el objetivo de adaptarla a las necesidades específicas del software utilizado para llevar a cabo la investigación:

La base de datos original contiene datos de fecha, hora y producción de energía fotovoltaica. Esta base viene especificada en intervalos de 15 minutos, pero para este estudio se han tenido en cuenta solamente intervalos de 1 hora. Este hecho ha supuesto por lo tanto la necesidad de seleccionar solamente aquellos registros que correspondan con un tramo horario.

Una vez filtrada la base de datos original, nos hemos quedado con una serie temporal, cuya temporalidad es horaria. Esta serie puede verse en la siguiente imagen:



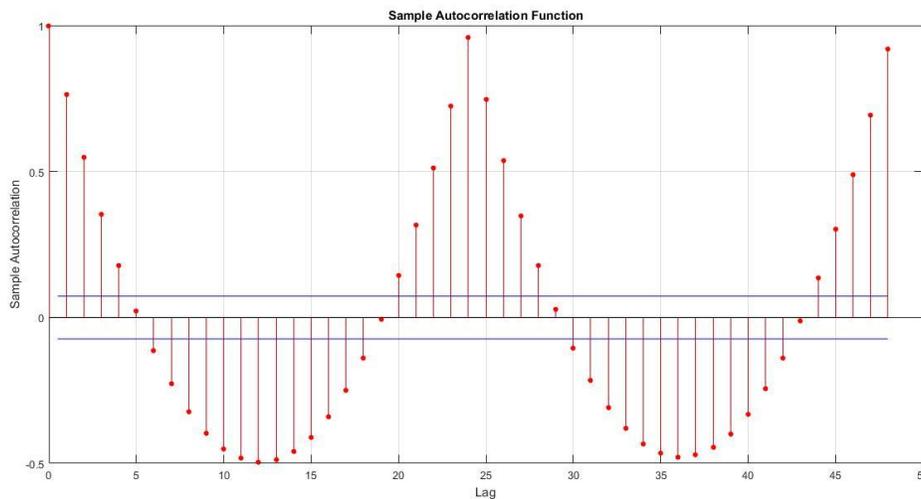
El corte en rojo que se aprecia en la imagen hace referencia a la separación de los datos en dos conjuntos: un conjunto de entrenamiento y un conjunto de validación.

Para esta aplicación se han considerado como datos de entrenamiento las 2 primeras semanas del mes de Octubre de 2017, mientras que las predicciones, una vez obtenido el valor óptimo del hiperparámetro ( $\gamma$ ) en el conjunto de entrenamiento, se han realizado para las 2 siguientes semanas. De esta forma, tenemos 672 datos en total, de los cuales 366 corresponden al conjunto de entrenamiento y los restantes 306 datos corresponden al conjunto de validación.

### 3.4 Elección del Regresor

Puesto que estamos tratando con una serie temporal, los elementos del regresor estarán compuestos por términos de la serie que corresponden a momentos pasados en el tiempo. Sin embargo, surge la duda de cuántos y cuáles componentes autorregresivos compondrán el regresor.

Para solucionar este dilema, podemos inspeccionar el autocorrelograma de la serie con el objetivo de detectar cuáles son los términos pasados que mayor autocorrelación presentan. A continuación se muestra dicho autocorrelograma para la serie de estudio.



Tal como puede observarse, existe una autocorrelación muy significativa en instante  $k - 1$ ,  $k - 2$ ,  $k - 24$ ,  $k - 25$  y  $k - 26$ . Partiendo de esta observación, tendremos por lo tanto que el regresor, para un instante del tiempo  $k$ , estará definido como:

$$r_k = [y_{k-1}, y_{k-2}, y_{k-24}, y_{k-25}, y_{k-26}]$$

### 3.5 Elección del gamma óptimo

El primer paso en el proceso de predicción es la obtención del valor óptimo para el hiperparámetro gamma ( $\gamma$ ). La función de este hiperparámetro es la de modificar el comportamiento de la función objetivo a minimizar, lo cual afecta directamente a la solución óptima que se obtiene para el vector

de pesos que se utilizarán para realizar las predicciones. Por lo tanto, el término de error que se obtenga de la obtención estará totalmente ligado a los valores de  $\gamma$  que se utilicen.

Para la obtención del  $\gamma$  óptimo, se ha utilizado el método Leave-One-Out, de forma que se han cogido todos los datos del conjunto de entrenamiento para generar sistemáticamente predicciones de cada uno de los valores de este conjunto. El procedimiento ha consistido en iterar sobre cada uno de los datos del conjunto y, para cada iteración:

1. Se dejan fuera de la muestra los datos del regresor y de la variable dependiente
2. Con los datos resultantes se resuelve el problema de optimización y se obtiene un vector de pesos óptimo
3. Con este vector de pesos se obtiene la predicción para el valor de la variable dependiente que se ha dejado fuera de la muestra

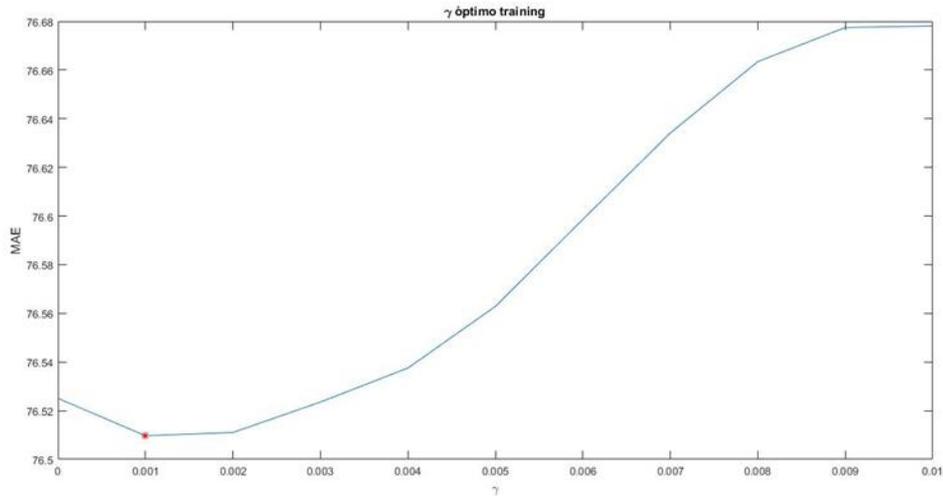
Para cada una de las predicciones del conjunto se calcula su el error de predicción como el valor absoluto de la diferencia entre el valor real y el valor predicho y una vez se obtienen todos los errores de predicción, se calcula una media de ellos que da como resultado el índice de error MAE (Mean Absolute Error).

El  $\gamma$  óptimo será aquel que consiga minimizar el error de predicción cometido, es decir, aquel que menor valor para el MAE presente.

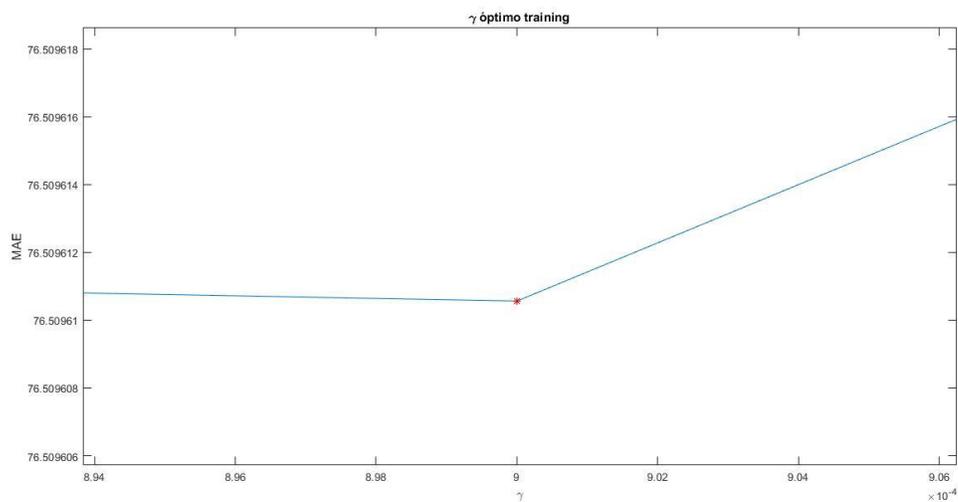
El proceso anteriormente descrito se ha realizado de forma sistemática para diferentes valores de gamma: en un primer momento se ha realizado una búsqueda superficial para valores de  $\gamma$  bastante alejados entre sí, para luego ir afinando cada vez más en las zonas donde se sospechaba que podía haber un descenso del término de error, hasta finalmente encontrar un valor óptimo para este hiperparámetro.

En un primer lugar se hizo una búsqueda para valores de  $\gamma$  comprendidos entre 0 y 1 con un espaciado de 0.25, donde se vio que la región de búsqueda era demasiado amplia y que el error parecía aumentar conforme el valor de gamma crecía. Posteriormente se afinó un poco más la

búsqueda cerca de valores más cercanos a 0, quedando la búsqueda encuadrada entre los valores 0 y 0.01, tal como se muestra en la siguiente gráfica:



Podemos observar que ya sí que se muestra un valor óptimo para el hiperparámetro gamma. Con el objetivo de obtener de forma más precisa el valor de gamma que minimiza el índice de error, volvemos a realizar una búsqueda, pero esta vez en valores comprendidos entre el 0.0008 y el 0.0015 con un tamaño de paso de 0.0001, ya que estos son los valores en los cuales parece estar incluido nuestro gamma óptimo, tal como se puede observar en la siguiente imagen:



Finalmente, obtenemos que el valor óptimo para el hiperparámetro gamma, en el conjunto de entrenamiento es de

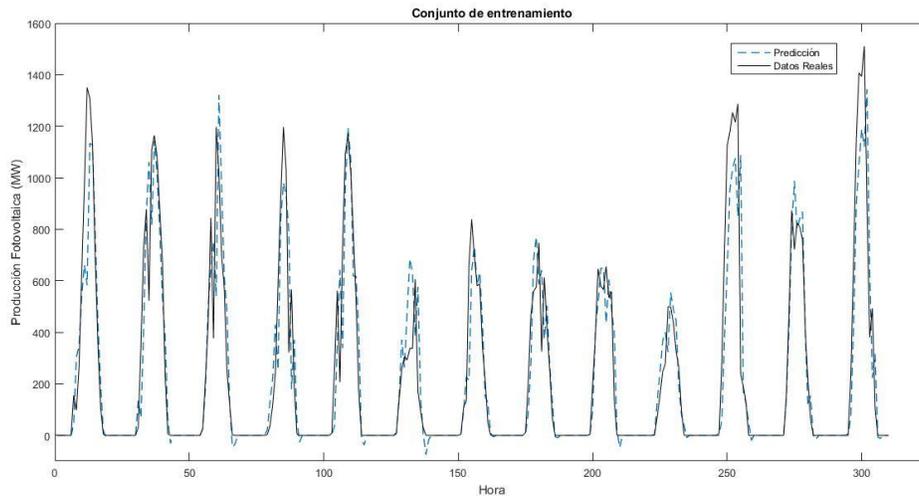
$$\gamma = 0.0009$$

El siguiente cuadro muestra los valores de gamma obtenidos y sus términos de error asociados, donde podemos ver que aparece nuestro gamma óptimo, el cual presenta el menor valor en el índice de error.

Tabla  $\gamma$  Óptimo

$\gamma$	MAE
0.0008	76.5096
0.0009	76.5096
0.0010	76.5097
0.0011	76.5098
0.0012	76.5099
0.0013	76.5101
0.0014	76.5100
0.0015	76.5097

Una vez obtenido el valor óptimo del hiperparámetro (gamma), inspeccionamos visualmente la calidad de las predicciones de nuestro predictor usando este hiperparámetro para la obtención del vector de pesos que nos servirá para realizar la predicción:



Como podemos observar, a priori, el comportamiento de nuestro predictor es bastante bueno, ya que parece ajustar muy bien a la serie temporal, pese a la alta volatilidad que presenta.

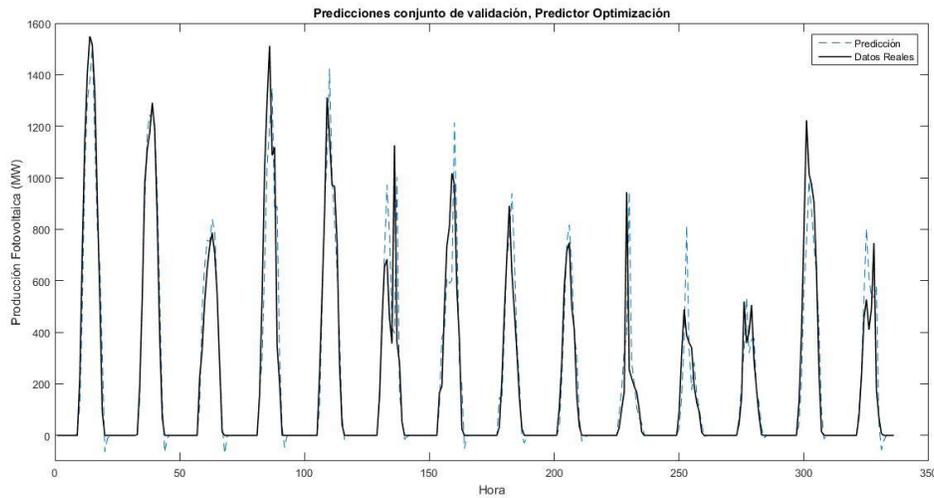
## 3.6 Generación de predicciones

Una vez ya hemos obtenido un valor óptimo para el hiperparámetro gamma y habiendo comprobado que, a priori, el predictor parece funcionar bastante bien en el conjunto de entrenamiento, resulta necesario contrastar la validez del predictor con datos no conocidos. Para ello, se realizan predicciones a un paso para cada uno de los datos del conjunto de validación, tal como a continuación se expondrá con mayor detalle.

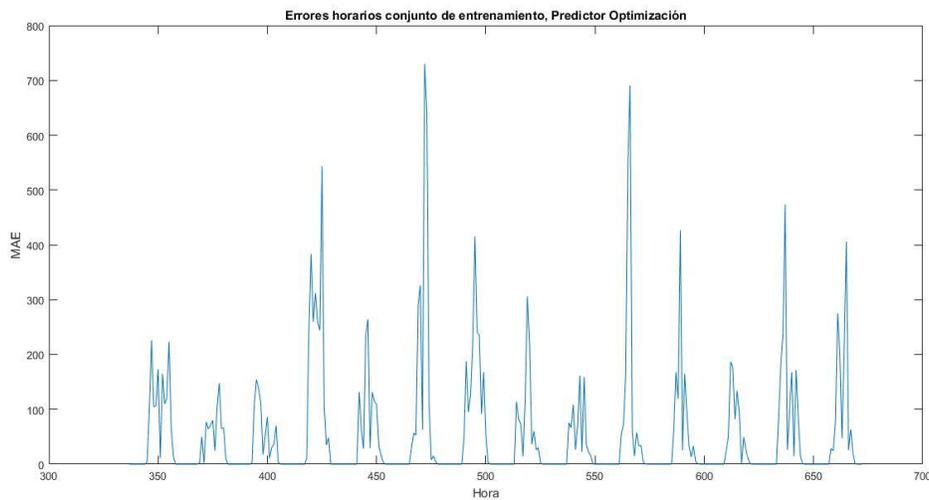
### 3.6.1 Predictor optimización

Para este predictor se utiliza el valor del hiperparámetro gamma obtenido en el conjunto de entrenamiento para resolver el problema de optimización propuesto en la presente investigación.

A continuación se muestra una gráfica que contiene los valores reales de la serie para el conjunto de validación, así como las predicciones realizadas con este predictor:



Tal como podemos observar en la gráfica anterior, efectivamente comprobamos que las predicciones son muy ajustadas a los valores reales de la serie en la mayoría de los casos. Con el objetivo de ver en qué puntos se comenten los mayores errores de predicción, a continuación se muestra la gráfica de errores horarios para cada una de las predicciones del período.

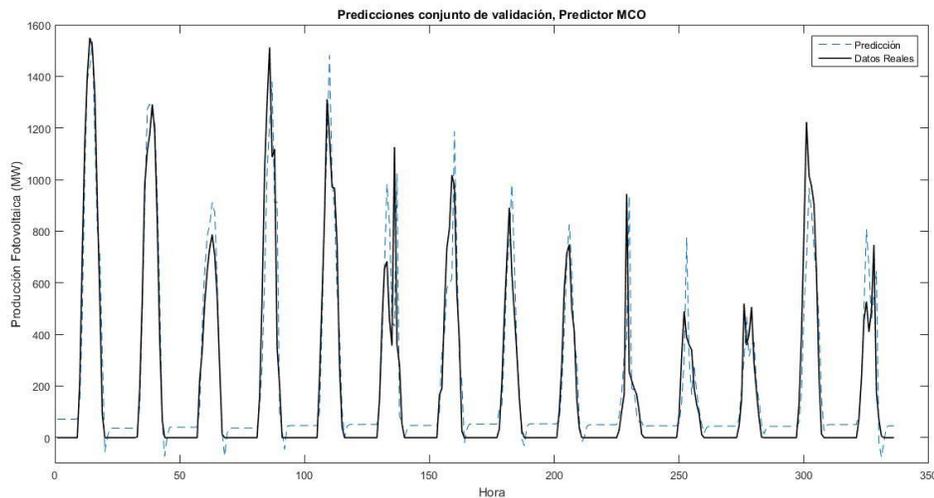


Podemos ver que los mayores errores se concentran sobre finales de la segunda semana y principios de la tercera, quedando un error medianamente notable a finales de la última semana.

El error medio o MAE asociado a la predicción es de **57.5511**.

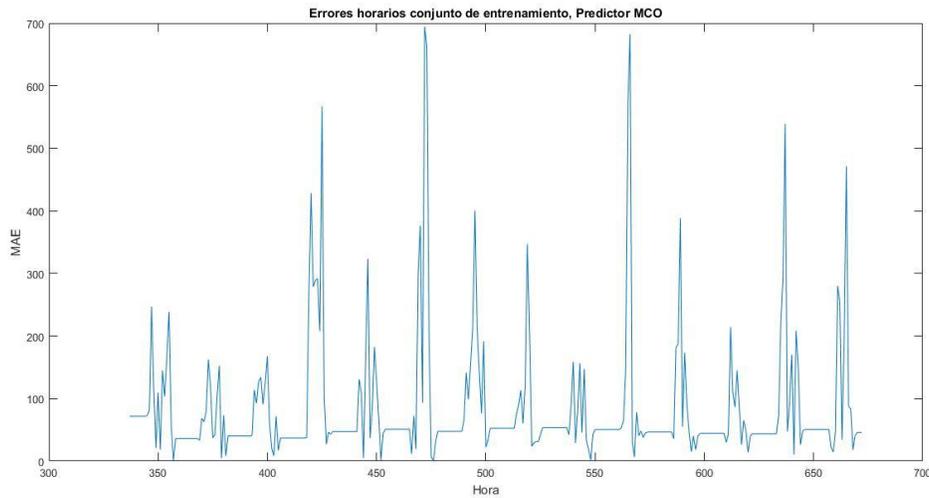
### 3.6.2 Predictor Mínimos Cuadrados Ordinarios

A continuación, se muestra la gráfica de la comparativa entre los datos reales del conjunto de validación y la predicción obtenida mediante el predictor basado en el método de regresión por Mínimos Cuadrados Ordinarios:



Como podemos observar, este predictor también es capaz de generar buenas predicciones, aunque, al contrario de lo que ocurría con el predictor anterior, este predictor parece que sobreestima en gran medida el valor del dato real. Esto puede verse en los picos, que son más altos y puntiagudos, pero también en las zonas donde el valor real de la producción de energía solar es 0, que coincide con las horas del día en las que no hay radiación directa del sol, donde el predictor da valores predichos mucho más altos que el valor real.

Se muestra también a continuación la gráfica de errores horarios para todo el período de predicción, que corresponde al conjunto de validación.



En esta gráfica podemos ver que los errores ahora son más altos y además observamos la tendencia de sobreestimación que se comentaba anteriormente.

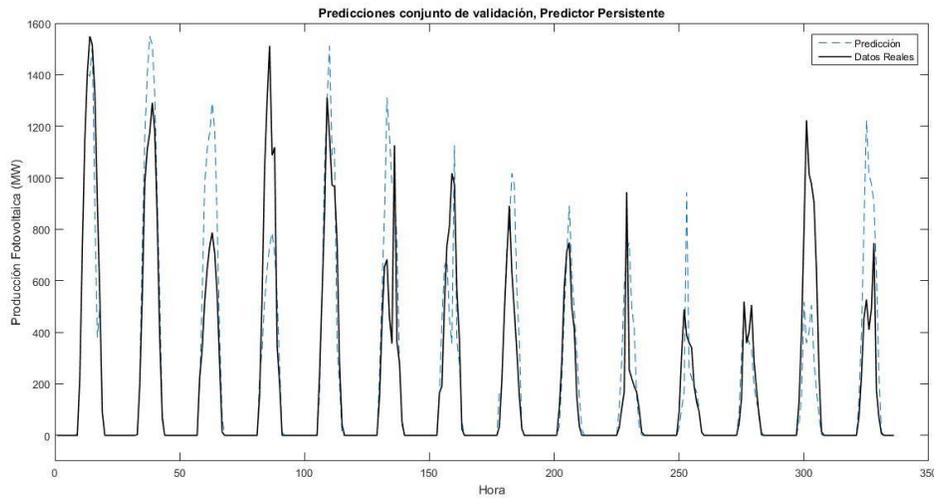
En términos medios, el MAE para este predictor se sitúa en **85.5691**.

### 3.6.3 Predictor persistente

El predictor persistente que se ha considerado en el presente estudio ha sido el siguiente:

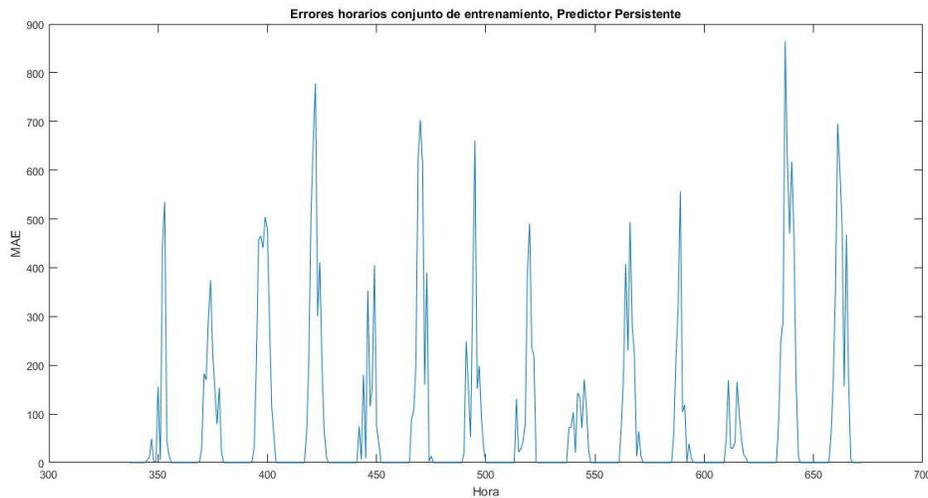
$$y_k = y_{k-24}$$

Con este predictor obtenemos los siguientes resultados:



En la anterior gráfica podemos observar que este predictor resulta ser el que peor predice, ya que debido a la gran volatilidad de los picos máximos diarios, al utilizar el persistente la información de la misma hora, pero del día anterior, en las zonas donde hay picos tiende a sobreestimar o subestimar en exceso.

Veamos la gráfica de errores asociados a las predicciones del predictor persistente.

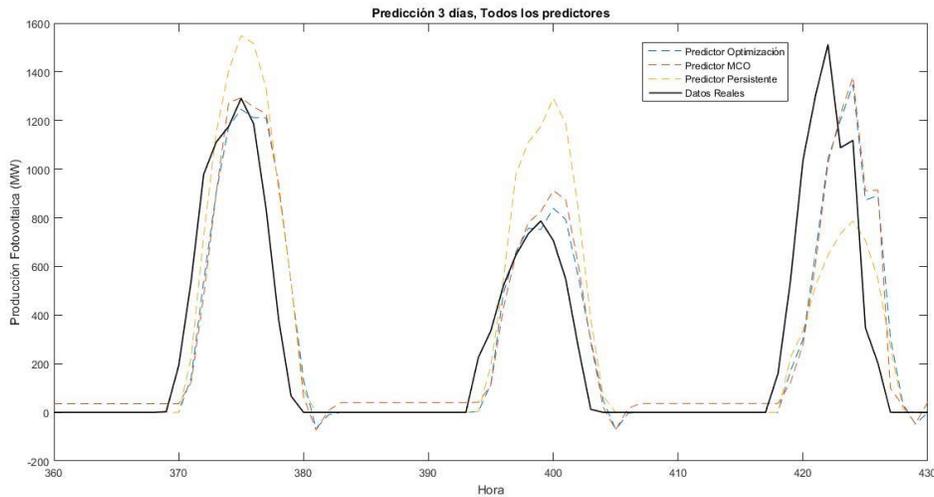


Efectivamente, podemos comprobar a simple vista cómo los errores han aumentado de media para este predictor.

Contrastamos este hecho con el MAE, ya que presenta un valor de **86.6661**, siendo este el término de error más alto para los 3 predictores.

### 3.7 Resultados

Para concluir este apartado práctico, a modo de comparativa global de los 3 predictores se muestra una gráfica para la predicción de 3 días con los datos reales y todos los predictores:



Podemos ver de forma notoria que el predictor que mejor ajusta a los datos reales es el predictor propuesto en nuestra investigación, seguido del predictor basado en Mínimos Cuadrados Ordinarios y, por último, el predictor persistente.

A modo de resumen, se muestra la tabla de errores medios para todo el período para estos 3 predictores.

Errores medios por modelo de predicción	
Predictor	Valor MAE
Optimización	<b>57.5511</b>
MCO	<b>85.5691</b>

Modelo ganador

Persistente	<b>86.6661</b>
-------------	----------------

## Capítulo 4 Conclusiones

De entre todos los predictores, vimos que el predictor de optimización, es el que obtiene las mejores predicciones. Y para estar seguro de la validez y el funcionamiento adecuado de este predictor, las predicciones se realizaron en un conjunto de validación. El resultado obtenido para este conjunto de validación confirma que el predictor en cuestión es apto para ser utilizado y hacer predicciones futuras de la producción de energía fotovoltaica.

Para futuros trabajos se propone comparar el método de optimización propuesto con otros métodos predictivos. Por otro lado, se requiere hacer un análisis más exhaustivo de los patrones subyacentes en los datos de una serie y para ello sería interesante incluir elementos exógenos como la fecha, estación del año, radiación solar y otras variables ambientales, como temperatura y luminosidad entre otros.

El método de predicción elaborado en esta investigación se ha realizado para el mercado de producción de energía fotovoltaica aplicado a Bélgica, este es un método genérico que sería extrapolable a cualquier problema de predicción similar.

## Referencias

- [1] Dincer, F. (2011). The analysis on photovoltaic electricity generation status, potential and policies of the leading countries in solar energy. *Renewable and Sustainable Energy Reviews*, 15(1), 713-720.
- [2] Elizbar A Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964.
- [3] Geoffrey S Watson. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A*, pages 359–372, 1964.
- [4] Roll, Jacob, Alexander Nazin, and Lennart Ljung. "Nonlinear system identification via direct weight optimization." *Automatica* 41.3 (2005): 475-490.
- [5] Gao, Jiti. *Nonlinear time series: semiparametric and nonparametric methods*. CRC Press, 2007.
- [6] Box, George EP, et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [7] Tjøstheim, Dag, and Bjørn H. Auestad. "Nonparametric identification of nonlinear time series: projections." *Journal of the American Statistical Association* 89.428 (1994): 1398-1409.
- [8] Härdle, Wolfgang. *Applied nonparametric regression*. No. 19. Cambridge university press, 1990.
- [9] Hamilton, James Douglas. *Time series analysis. Vol. 2*. Princeton: Princeton university press, 1994.
- [10] Haggan, Valérie, and Tohru Ozaki. "Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model." *Biometrika* 68.1 (1981): 189-196.

- [11] Chan, Kung Sik, and Howell Tong. "On estimating thresholds in autoregressive models." *Journal of time series analysis* 7.3 (1986): 179-190.
- [12] Tong, Howell. *Threshold models in non-linear time series analysis*. Vol. 21. Springer Science & Business Media, 2012.
- [13] Milanese, Mario. "Worst-Case  $l_1$  Identification." *Bounding Approaches to System Identification*. Springer US, 1996. 169-181.
- [14] Milanese, Mario, and Carlo Novara. "Set membership identification of nonlinear systems." *Automatica* 40.6 (2004): 957-975.
- [15] Boyd, Stephen, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [16] Bravo, J. M., et al. "A General Framework for Predictors Based on Bounding Techniques and Local Approximation." *IEEE Transactions on Automatic Control* 62.7 (2017): 3430-3435.