

Predicción de consumos eléctricos

Carles Figuera Penedo

Máster universitario en Ciencias de Datos

TFM – Área 5

Nombre Consultor/a: Sergio Trilles Oliver

Profesor/a responsable de la asignatura: Albert Solé Ribalta

Junio de 2020



Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

FICHA DEL TRABAJO FINAL

Título del trabajo:	<i>Predicción de consumos eléctricos</i>
Nombre del autor:	<i>Carles Figuera Penedo</i>
Nombre del consultor/a:	<i>Sergio Trilles Oliver</i>
Nombre del PRA:	<i>Albert Solé Ribalta</i>
Fecha de entrega (mm/aaaa):	06/2020
Titulación::	<i>Máster universitario en Ciencias de Datos</i>
Área del Trabajo Final:	<i>M2.982 - TFM - Área 5</i>
Idioma del trabajo:	<i>Castellano</i>
Palabras clave	<i>energy consumption prediction, time series forecasting, deep learning.</i>
Resumen del Trabajo (máximo 250 palabras): <i>Con la finalidad, contexto de aplicación, metodología, resultados y conclusiones del trabajo.</i>	
<p>Hoy en día la energía eléctrica es un recurso esencial para la mayoría de la población. Dispositivos electrónicos, calefacción, electrodomésticos, iluminación, comunicaciones y a día de hoy muchos vehículos de transporte, son claros ejemplos de que la importancia del abastecimiento eléctrico para la población.</p> <p>Las distribuidoras eléctricas deben dar servicio a la población de un recurso tan importante con dos importantes problemáticas: es un recurso difícil de almacenar (en gran volumen) y difícil de producir de forma inmediata. Así que se debe generar un volumen de energía en equilibrio con el volumen de demanda, de ahí viene la necesidad de predecir constantemente cuál será el consumo de la población.</p> <p>Aunque existen ciertos patrones que nos ayudan a predecir el comportamiento consumista de la población, existen otros factores como la climatología o la economía que pueden modificar dichos patrones y provocar un error en la predicción de consumo, dejando en consecuencia a parte de la población sin abastecimiento eléctrico o produciendo más energía de la demandada.</p> <p>Para predecir estas necesidades de consumo, existen distintos algoritmos y modelos de predicción que nos ayudan a solventar este problema. En este proyecto trabajaremos con un <i>dataset</i> de mediciones de consumo (<i>time series</i>) para investigar y comparar las predicciones obtenidas de los distintos modelos de <i>deep learning</i> estudiados en el proyecto que se ajusten a esta tipo de problemática o necesidad.</p>	

Abstract (in English, 250 words or less):

Nowadays the electric energy is an essential resource for most of the population. Electronic devices, heating, appliances, lighting, communications and today many transport vehicles, are clear examples of the importance of electricity supply for the population.

Electricity distributors must supply the population of such an important resource with two important problems: it's a resource difficult to store (in large volumes) and difficult to produce immediately. So a volume of energy must be generated in equilibrium with the volume of demand, hence the need to constantly predict what the population's consumption will be.

Although there are certain patterns that help us predict the consumer behavior of the population, there are other factors such as the weather or the economy that can modify these patterns and cause an error in the prediction of consumption, leaving consequently part of the population without supply electric or producing more energy than the defendant.

To predict these consumption needs, there are different algorithms and prediction models that help us solve this problem. In this project we will work with a dataset of consumption measurements (time series) to investigate and compare the predictions obtained from the different deep learning models studied in the project that fit this type of problem.

Índice

M2.982 - TFM - Área 5.....	i
1. Introducción.....	4
1.1 Contexto y justificación del Trabajo	4
1.2 Motivación personal.....	7
1.3 Objetivos del Trabajo.....	7
1.4 Enfoque y método seguido.....	8
1.5 Planificación del Trabajo.....	8
1.6 Breve sumario de productos obtenidos.....	9
1.7 Breve descripción de los otros capítulos de la memoria.....	9
2. Estado del arte.....	10
2.1 Series temporales.....	10
2.1.1 Estacionariedad.....	10
2.1.2 Linealidad.....	11
2.1.3 Tendencia.....	11
2.1.4 Estacionalidad.....	12
2.2 Componentes de una serie temporal.....	12
2.3 Clasificación de una serie temporal.....	14
2.4 Predicción de series temporales.....	15
2.5 Modelo ARIMA.....	16
2.6 Modelos ANN.....	17
2.6.1 Red neuronal CNN.....	20
2.6.2 Red neuronal LSTM.....	21
5. Bibliografía.....	23

1. Introducción

1.1 Contexto y justificación del Trabajo

La electricidad es un recurso muy importante para la población. Son muchas y muy variadas las necesidades que la población cubre con la energía eléctrica (calefacción, entretenimiento, iluminación, comunicación, transporte,...). Los avances en tecnología hacen que cada día haya más demanda de este recurso, y por ello, la disponibilidad de energía eléctrica es tan importante para hogares, empresas, hospitales, escuelas, etc. Dada su importancia, se podría decir que la disponibilidad de electricidad es un claro indicador de crecimiento y desarrollo económico para una población.

Pero cuando hablamos en términos de disponibilidad, la energía eléctrica muestra dos grandes problemas para las empresas distribuidoras.

En primer lugar, la electricidad es un recurso muy costoso de almacenar en grandes volúmenes. Otros recursos como el petróleo o el gas, se pueden almacenar en depósitos a la espera de ser usados o consumidos. Pero la electricidad tiene unas características que hacen que sea muy costoso de almacenar, aunque hay empresas como *Endesa* o *Iberdrola* que empiezan a trabajar en esta línea de investigación usando baterías de gran tamaño [1][2]. Pero se sabe que las baterías pierden eficiencia con el tiempo y uso, haciendo que por el momento, sea muy costoso el almacenaje de energía. Por eso decimos que el consumo de energía eléctrica es un consumo en tiempo real, ya que se consume a la vez que se genera, sin posibilidad de almacenar sobrantes. Así que la generación y su consumo deben estar en equilibrio.

En segundo lugar, la producción de energía eléctrica no es inmediata. Es decir, que si se dispara el consumo eléctrico en un intervalo de tiempo corto, las distribuidoras tienen muy poco tiempo de reacción para producir más energía. Aunque la respuesta de las centrales eléctricas tiene que ser rápida para no desabastecer parte de la población, muchas veces, sólo el hecho de incrementar la producción eléctrica puede costar entre minutos y horas, tiempo necesario para por ejemplo, abrir compuertas en una central hidroeléctrica o incrementar la cantidad de combustible de las centrales térmicas. Además hay que tener en cuenta que normalmente las fuentes de energía eléctrica con capacidad de respuesta más rápida acostumbran a ser las más contaminantes (centrales termoeléctricas de carbón, gas o gasóleo). La eficiencia de las centrales renovables dependen de recursos naturales como el viento, el caudal hidrológico o la luz solar, de las cuáles no tenemos control.

En resumen, el coste elevado de almacenar la energía y el coste económico/medioambiental de producirla de forma rápida, hacen que sea de gran valor y un tema relevante la posibilidad de realizar predicciones de consumo eléctrico de los usuarios de la red eléctrica. A continuación podemos observar un ejemplo de la predicción en tiempo real [3] de la que dispone la REE (Red Eléctrica Española).

Podemos observar como en un mismo día, la demanda de energía eléctrica (línea amarilla) es muy irregular. Desciende por las noches cuando la población descansa y presenta dos picos acentuados a primera hora de la mañana (cuando la población y las empresas empiezan su actividad diaria) y al atardecer (cuando la población vuelve a sus hogares).

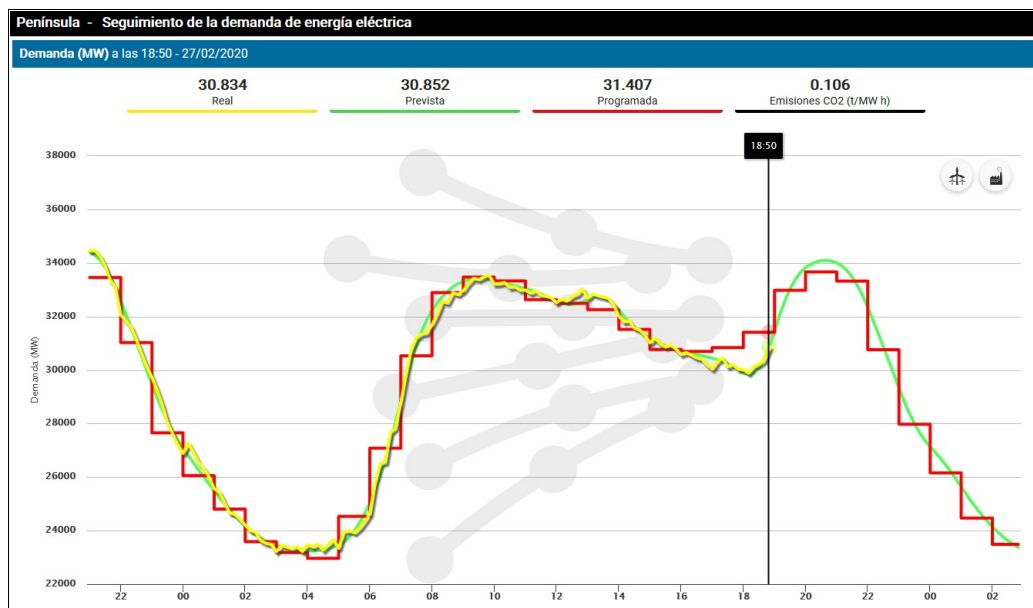


Figura 1.1. Demanda y predicción de energía eléctrica por la REE

Esta gráfica muestra cómo las predicciones de consumo pueden variar según las horas del día, aunque a la vez también nos muestra que existen ciertos patrones. Pero no sólo pueden variar según la hora, sino que hay otras características o factores a tener en cuenta cuando hablamos de predicción.

A continuación se muestra un listado de todas ellas:

1. Horas del día: como ya hemos comentado, el consumo es diferente según horas del día (figura 1.1).
2. Días de la semana: los consumos eléctricos son distintos en días laborables o en fin de semana. El fin de semana acostumbra a bajar el consumo ya que muchas empresas cesan su actividad.
3. Estaciones del año: por ejemplo, el uso de calefacción en invierno y aires acondicionados en verano hace que se tenga en cuenta la época del año para predecir de forma más exacta el consumo eléctrico.
4. Eventos puntuales: una huelga general provoca la bajada del consumo de forma notoria. O bien la noche de fin de año puede que hacer que se incremente la demanda ya que los locales de ocio trabajan más horas.
5. Meteorología: existe una correlación entre el consumo de energía eléctrica y la temperatura ambiente. Así que también deberá existir una relación entre la predicción de consumo eléctrico y la predicción meteorológica.
6. Economía: la salud financiera de una zona geográfica también aplica al consumo eléctrico. Por ejemplo, durante la última crisis económica, el cierre de muchas fábricas y empresas hizo disminuir el consumo eléctrico de forma notoria.

Así que podemos hablar de dos grandes tipos de factores que afectan al consumo de electricidad: el calendario y la climatología. Estos factores tienen patrones conocidos. Por una parte, la temperatura y las lluvias siguen un patrón anual conocido (figura 1.2), y por otra parte, el factor temporal o de calendario (días festivos, días laborables, horas diurnas, horas nocturnas,...) también sigue otro patrón conocido (figura 1.3). Así que podemos trabajar las predicciones de consumos eléctricos como si fueran predicciones de series temporales.

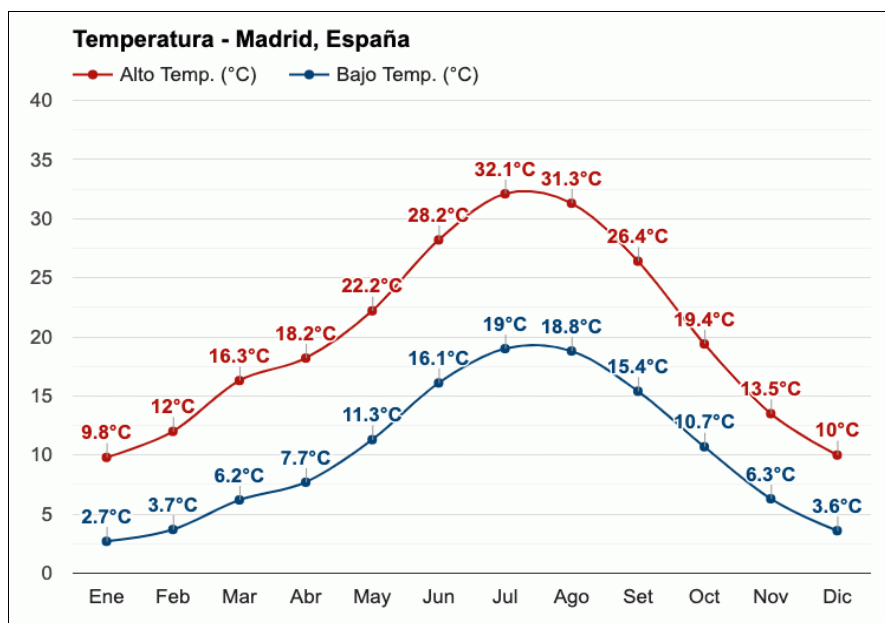


Figura 1.2. Promedio anual de temperaturas en Madrid [4]

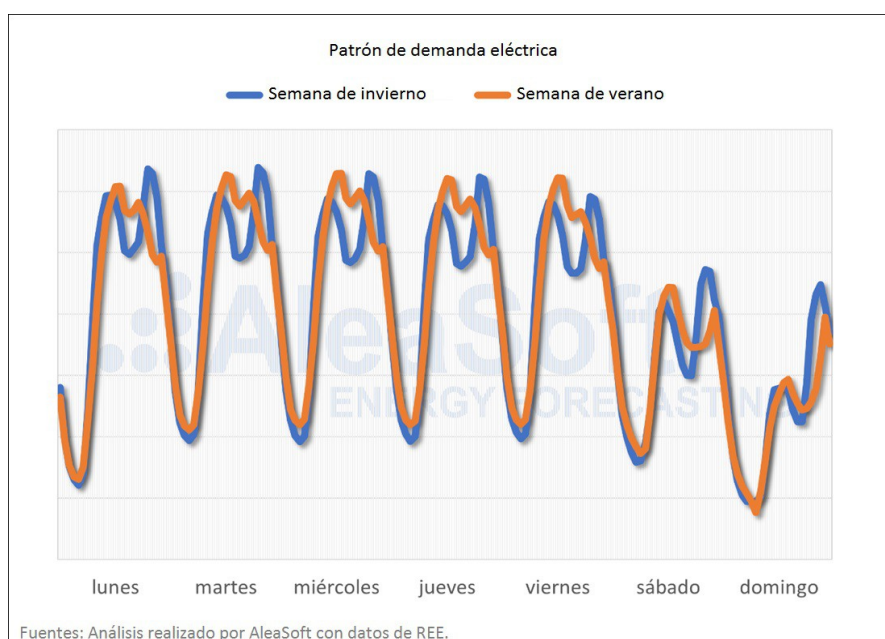


Figura 1.3. Demanda eléctrica media en España 2014-2018 según semana y estación del año

En este trabajo estudiaremos y compararemos diferentes modelos de predicción que encajan para resolver problemas del tipo series temporales como el introducido en este apartado.

1.2 Motivación personal

Creo que de todas las disciplinas del *Big Data*, las de *machine learning* y *deep learning* es la más interesante. La idea de predecir, averiguar qué va a suceder antes de que suceda de forma automática, solamente teniendo en cuenta datos de lo que ha pasado anteriormente me fascina.

Además, al licenciarme en 2011, mi primer proyecto profesional fue en un proyecto de *Smartcity* en Málaga [5], donde comprendí las dificultades que tenía administrar la energía eléctrica de toda una zona de la ciudad. El proyecto fue construir una prueba piloto de introducción de sistemas *Big Data* en la ciudadanía, creando lo que llamaban una 'ciudad inteligente'. Fue una gran experiencia, en mi departamento se trabajaba mucho en la monitorización del uso de la energía eléctrica para el ahorro energético.

Así que este trabajo trata sobre dos temas que me fascinan y me aportará mucho a nivel personal.

1.3 Objetivos del Trabajo

A continuación se muestra un listado de objetivos que se quieren realizar en este trabajo de fin de máster:

- Entender la necesidad y problemática de la predicción de consumos eléctricos.
- Estudiar el estado del arte actual de algoritmos estadísticos o de aprendizaje automático (*machine learning*) y profundo (*deep learning*) para predicción de series temporales (*time series forecasting*).
- Tomar un *dataset* de series temporales sobre consumos eléctricos, analizarlo y prepararlo para la creación de modelos predictivos.
- Crear y entrenar distintos modelos de predicción a partir del *dataset* mencionado.
- Preprocesar el *dataset* anterior para poder entrenar y evaluar de forma satisfactoria los distintos modelos predictivos creados.
- Analizar y comparar los distintos resultados obtenidos con los modelos anteriores mediante un *dataset* de evaluación.
- Evaluar posibles mejoras en los modelos creados en el estudio (por ejemplo: *Dropout* en modelos DL).
- Estudiar posibles líneas futuras de estudio (por ejemplo: modelos híbridos, etc).

1.4 Enfoque y método seguido

La estrategia a seguir en este proyecto será la de investigar y comparar los modelos predictivos usados en la actualidad, productos ya existentes. Definiremos su funcionamiento, crearemos modelos de test, los entrenaremos y los compararemos tanto en resultados como en rendimiento. También intentaremos encontrar solución a los problemas o errores encontrados en dichos modelos.

La metodología a seguir, por las características de la investigación y el número de participantes, será la de un proceso de desarrollo secuencia, es decir, desarrollo en cascada. Puede ser que en la fase de creación de modelos predictivos se pase a una metodología más *agile* del tipo *scrum* si pensamos en cada modelo predictivo a reproducir en el estudio como en un entregable.

Dichas investigaciones o experimentos serán a futuro reproducibles, ya que se almacenaran, además de los productos resultantes de la investigación (resultados, figuras, tablas, memoria, etc), los códigos fuente, los *datasets* o datos originales y los ya preprocesados, junto con las instrucciones a seguir para su repetición futura del estudio en caso de ser necesario.

Todo producto resultante se entregará en el siguiente repositorio GitHub:

<https://github.com/cfiguerap/uoc-data-science-tfm>

1.5 Planificación del Trabajo

A continuación, se muestra un diagrama de *Gantt* de la planificación temporal inicial que se ha establecido para el proyecto:





Figura 1.4. Diagrama de *Gantt* con planificación temporal del proyecto

1.6 Breve resumen de productos obtenidos

El producto entregable del presente proyecto es una comparativa de rendimiento y resultados de los modelos predictivos más comunes de predicción de series temporales basados en un *dataset* de consumo eléctrico.

Además, se hará entrega de la presente memoria y resto de documentación, junto los ejecutables producto del estudio realizado, en el siguiente repositorio *GitHub*: <https://github.com/cfiguerap/uoc-data-science-tfm>

1.7 Breve descripción de los otros capítulos de la memoria

El siguiente esquema muestra la estructura del presente documento, dividiéndose en los siguientes capítulos:

- Capítulo 2. Estudio del estado del arte: introducción a modelos predictivos usados para la predicción de consumos u otros tipos de series temporales.
- Capítulo 3. Análisis y preprocesado del *dataset* del caso de estudio: análisis exhaustivo del *dataset* escogido para el caso de estudio y preprocesado, en caso de ser necesario, para la construcción de modelos predictivos.
- Capítulo 4. Creación y entrenamiento de los modelos predictivos: creación de diferentes modelos predictivos a partir de los datos del *dataset* antes mencionado.
- Capítulo 5. Resultados obtenidos de los modelos predictivos: recopilación de datos y resultados obtenidos tras la evaluación de dichos modelos predictivos.
- Capítulo 6. Conclusiones del estudio realizado: comparativa de los diferentes resultados obtenidos en el capítulo anterior.
- Capítulo 7. Futuras líneas de investigación: introducción a futuras líneas de investigación sobre los modelos predictivos de series temporales.

2. Estado del arte

2.1 Series temporales

Una serie temporal es una secuencia de valores discretos ordenados en el tiempo. Normalmente, en los sectores industriales o de ingeniería, son el resultado de una monitorización de un proceso con cierto grado de interés. Por ejemplo, en el caso de estudio de este proyecto, el consumo eléctrico que estudiaremos es una serie temporal resultante de la continua monitorización de las distribuidoras eléctricas.

$$X = \{X_1, X_2, \dots\} \text{ o } \{X_k\}_{k \geq 1}$$

Figura 2.1. Notación de una serie temporal

Las series temporales pueden ser regulares o irregulares. Las regulares tienen una distancia temporal constante entre sus valores (diarios, semanales,...), mientras que las irregulares tienen una distancia variable

Las series temporales son muy importantes para la predicción, para poder predecir correctamente los consumos futuros, necesitamos conocer los consumos del pasado. Pero para poder predecir con menor error, no sólo necesitamos los valores de interés (el consumo), también necesitamos conocer otros factores relacionados con el consumo (meteorológicos, temporales, económicos, etc).

Las series temporales tienen una serie de características o propiedades [6]. Cada serie puede mostrar alguna o más de una de estas propiedades:

2.1.1 Estacionariedad

Esta propiedad está relacionada con la media y varianza de los valores de la serie temporal. Se dice que es estacionaria cuando la media y la varianza se mantienen constantes con el tiempo y además no presentan una tendencia.

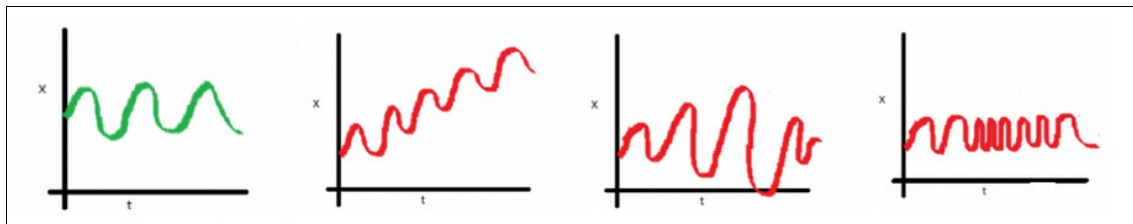


Figura 2.2. En verde, ejemplo de una serie temporal estacionaria y en rojo distintos ejemplos de series no estacionarias [7]

2.1.2 Linealidad

La linealidad de una serie temporal indica que la forma que tiene depende del tiempo, es decir, los valores de la serie se pueden determinar mediante el tiempo con una función lineal. Si la serie es lineal, entonces puede ser representada por una función o ecuación lineal. En cambio, las no lineales sólo pueden ser representadas mediante ecuaciones no lineales.

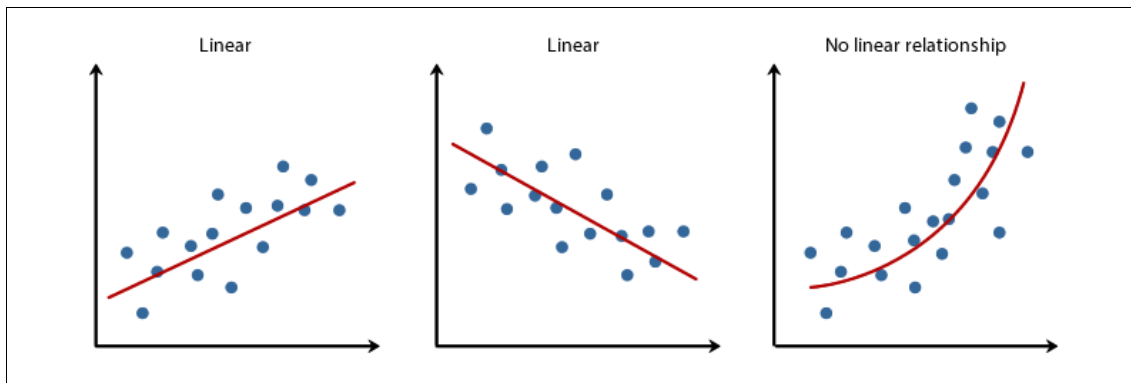


Figura 2.3. Ejemplos de series temporales lineales y no lineales

2.1.3 Tendencia

El componente de tendencia hace referencia al movimiento general de largo plazo de la serie temporal. Es decir, la tendencia describe el comportamiento de la serie a largo plazo. El análisis de tendencias es importante en el pronóstico de series de tiempo. En la práctica, se usan técnicas de regresión lineal y no lineal que ayudan a identificar el componente de tendencia no monótono en la serie temporal.

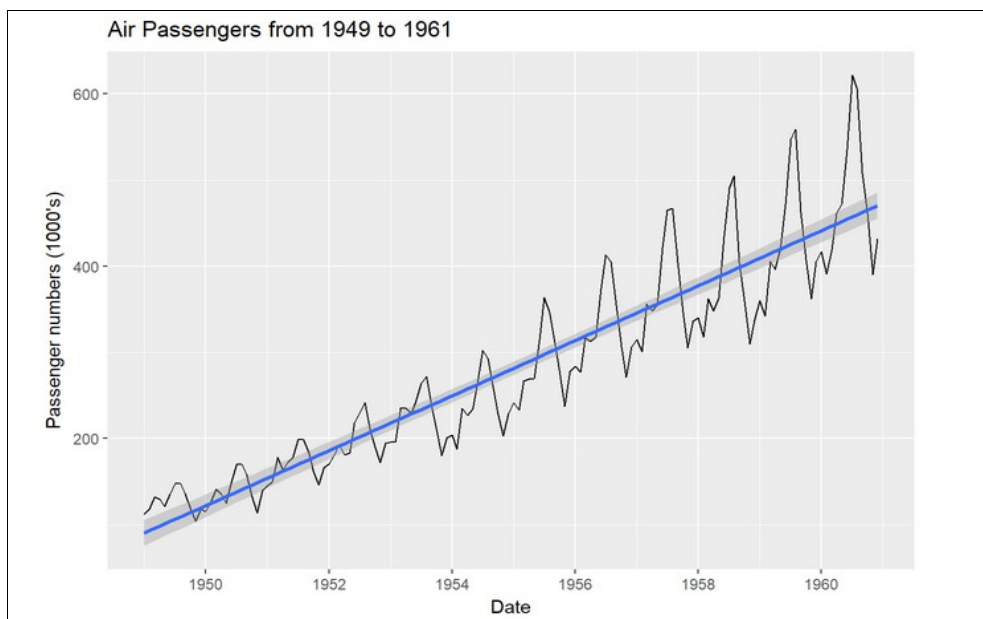


Figura 2.4. Ejemplo de serie temporal con distintas tendencias

2.1.4 Estacionalidad

Los componentes estacionales se pueden definir como oscilaciones que se producen en períodos iguales o inferiores a un año y que se reproducen de manera reconocible en los diferentes años. En otras palabras, hace referencia a cierta periodicidad de la serie o variación de cierto período, ya sea anual, semestral, trimestral, mensual, etc.

Normalmente, el objetivo principal del análisis de series temporales estacionales se centra en la detección del carácter de su fluctuaciones periódicas y sobre su interpretación. Los ejemplos típicos de series temporales estacionales, como en nuestro caso de estudio, son distribuciones de energía, gas, agua u otros sistemas, donde la predicción de las demandas de los consumidores representa el problema básico.

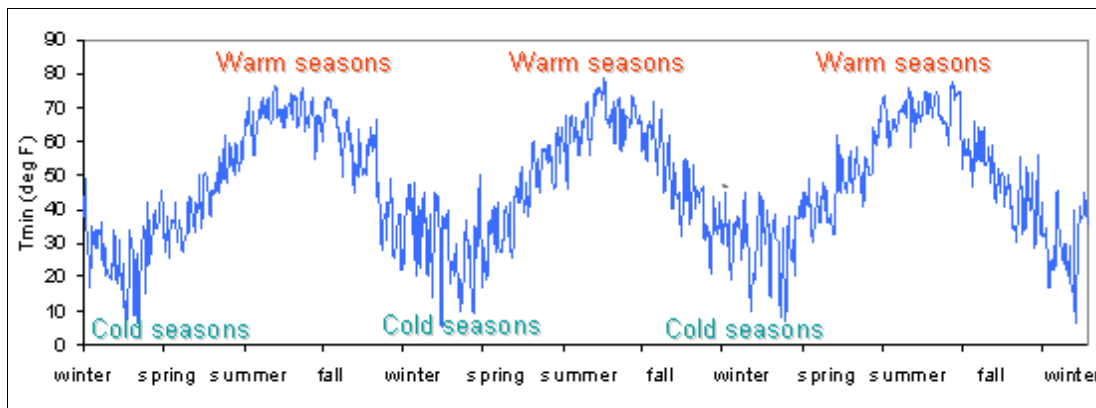


Figura 2.5. Serie temporal estacional, podemos apreciar como el período oscila de forma anual.

2.2 Componentes de una serie temporal

Las series temporales se pueden descomponer en cuatro componentes diferenciados que son: tendencia, estacionalidad, ciclo y ruido [8]. Cuando hablamos de análisis de series temporales, dichos componentes, en combinación los unos con los otros, definen el comportamiento de la serie temporal. A continuación definimos los componentes [8]:

- Tendencia: movimiento general que describe el comportamiento de la serie temporal a largo plazo.
- Ciclo: fluctuaciones alrededor de la tendencia que se repiten de forma más o menos periódica y de amplitud superior al año.
- Estacionalidad: oscilaciones que se producen en períodos iguales o inferiores a un año (diaria, semanal, mensual, trimestral, etc).
- Ruido: oscilaciones que no responden a ningún patrón de comportamiento conocido. Suelen ser debidos a factores fortuitos.

Las series temporales pueden llegar a representar una variedad de patrones infinitos. Y si se descompone la serie en estos distintos componentes que la forman, podemos llegar a descubrir distintos patrones en los componentes. Este hecho, nos permite poder conocer el comportamiento de las series temporales de una forma más clara, ya que estudiando cada componente por separado nos aporta información más útil y más fácil de conseguir que estudiando directamente la serie temporal original.

Por ejemplo, cuando analizamos el volumen de pedidos de una compañía para predecir sus futuras ventas, podemos analizar por separado cada componente de la serie temporal de ventas anteriores. En este caso, la tendencia nos ofrecería información valiosa sobre el ritmo de ventas sin vernos afectados por las fluctuaciones derivadas de los meses de vacaciones o Navidad. Mientras que la estacionalidad nos daría una idea de cuáles son los mejores y peores meses del año de cara a las ventas de la compañía.

En la práctica, la distinción entre tendencia y ciclo es problemática ya que, cuando se tienen pocas observaciones, es fácil confundir estos dos componentes. Así que cuando descomponemos una serie temporal en sus componentes, normalmente se combina la tendencia y el ciclo en un solo componente. De esta forma, podemos pensar que una serie temporal consta de sólo tres componentes: un componente de tendencia, un componente estacional y un componente de ruido [10].

En la siguiente figura (2.6) podemos observar los tres distintos componentes que forman una serie temporal:

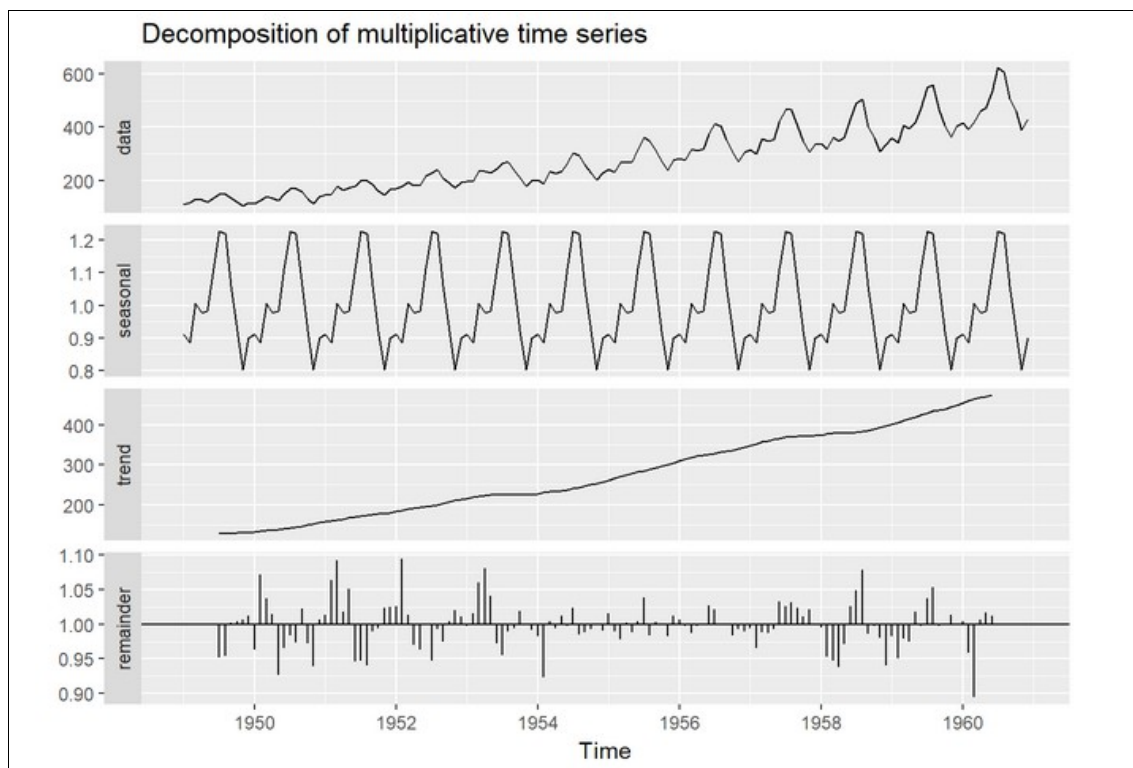


Figura 2.6. Ejemplo de descomposición de serie temporal.

2.3 Clasificación de una serie temporal

Como hemos visto en los apartados anteriores, dependiendo de los valores que forman las series temporales, las podemos clasificar en:

- Estacionarias o no estacionarias
- Estacionales o no estacionales
- Lineales o no lineales
- Caóticas (valores aleatorios sin periodicidad)

Pero además existen otros dos niveles más de clasificación [6]:

- Univariantes o Multivariantes:

Los valores de las series univariantes sólo dependen del tiempo (unidimensional). Por ejemplo, datos recopilados de un sensor que mide temperatura.

Las series multivariantes dependen del tiempo y otros factores (multidimensional). Por ejemplo, en nuestro caso de estudio, el consumo eléctrico depende del factor temporal pero también de la meteorología.

Las multivariantes, al depender de más factores, tienden a tener por probabilidad más *outliers*. Además, normalmente estos distintos factores pueden presentar una correlación que da forma al patrón de la serie temporal.

- Aditivas o Multiplicativas:

Las series aditivas son las formadas por la suma de sus componentes (tendencia, estacionalidad y ruido).

$$x_t = T_t + E_t + C_t + R_t$$

Figura 2.7. Ecuación de una serie temporal aditiva (tendencia, estacionalidad, ciclo y residuo)

En cambio, las multiplicativas son las formadas por el producto de sus componentes.

$$x_t = T_t \cdot E_t \cdot C_t \cdot R_t$$

Figura 2.8. Ecuación de una serie temporal multiplicativa (tendencia, estacionalidad, ciclo y residuo)

2.4 Predicción de series temporales

La predicción de series temporales es un tema que hoy en día interesa mucho en distintos sectores: financiero, comercial, energético, sanitario,... Desde predicciones bursátiles (IBEX35) hasta predicciones de contagios virales (COVID19), pasando por la predicción de demanda energética como sucede en nuestro caso de estudio.

El propósito de la predicción de series temporales suele ser doble: comprender las causas que dan lugar a una serie temporal (análisis) y predecir sus valores futuros a partir de sus valores pasados o conocidos (predicción).

Existen distintas técnicas de pronóstico o predicción de series temporales, algunas de ellas más modernas que otras. Existen desde técnicas puramente estadísticas (AR, MA o ARIMA) hasta técnicas más modernas de *Deep Learning* (ANN o SVM). Pero también existen casos de modelos o técnicas híbridas donde se usan más de un tipo de técnicas conjuntamente.

A continuación se introducen estos dos grupos importantes mencionados:

1) Técnicas tradicionales

Entre las distintas técnicas tradicionales destacan los modelos estadísticos, como el modelo autoregresivo (AR) o promedio móvil (MA), y en particular el modelo ARIMA (combinación de los dos modelos anteriores) que ha demostrado un mejor rendimiento en precisión y exactitud entre las predicciones de estas técnicas tradicionales.

Pero este tipo de modelos presentan una serie de desventajas o limitaciones a la hora de modelar problemas reales [11]:

- Requieren datos completos: valores perdidos u *outliers* pueden afectar considerablemente a la predicción resultante.
- Se basan en relaciones lineales: sus suposiciones tienen una base lineal que empeoran la predicción con series temporales complejas no lineales.
- Se basan en series univariadas: no se obtiene una buena predicción cuando los valores a predecir dependen de más de un *input* o entrada.
- Pronostican a corto tiempo: por lo general, funcionan bien en predicciones a corto plazo pero no a largo.

2) Técnicas de aprendizaje automático

Con el avance en los últimos años en tecnología y desarrollo de algoritmos de aprendizaje automático, se han desarrollado nuevos modelos que parece que obtienen mejores predicciones que los modelos tradicionales. Entre ellos destacan dos modelos de aprendizaje supervisado con redes neuronales: las LSTM (*long short-term memory*) y CNN (*convolutional neural network*).

Este tipo de técnicas tienen una serie de ventajas frente a los modelos tradicionales:

- Capacidad para aproximar funciones no lineales.
- Capacidad para tratar el ruido y *outliers*.
- Capacidad para aceptar entradas multivariantes.
- Capacidad para realizar pronósticos de varios pasos.

En los siguientes apartados vamos a entrar más en detalle de los modelos más usados en la actualidad para la predicción de series temporales. Se pueden usar muchos modelos distintos pero nos centraremos en los más usados.

2.5 Modelo ARIMA

El modelo ARIMA (*Autoregressive Integrated Moving Average*) es un método que se utiliza tanto para el análisis como la predicción de series temporales. Es un tipo de modelo estadístico que nos da una ecuación sobre una serie temporal en función de sus propios valores pasados, de modo que dicha ecuación se puede utilizar para pronosticar valores futuros. Como ya hemos dicho antes, es una combinación de los modelos autoregresivos (AR), promedio móvil (MA). Sus siglas se forman por los siguientes componentes:

- AR (Autoregresión): Un modelo que usa la relación dependiente entre una observación y un número de observaciones pasadas.
- I (Integrado): El uso de diferenciación entre los valores para ayudar a hacer la serie estacionaria. La diferenciación es un proceso que nos permite eliminar la tendencia de una serie temporal restando a cada valor de la serie el valor anterior.
- MA (Promedio móvil): Un modelo que usa la dependencia entre un valor y un error residual de un modelo de promedio móvil aplicado a valores anteriores.

Cada una de estos componentes forma parte en el modelo como un parámetro. Se utiliza una notación estándar de ARIMA (p, d, q) donde los parámetros se sustituyen por valores enteros para indicar un modelo ARIMA específico.

Los parámetros del modelo ARIMA se definen de la siguiente manera:

- P**: Número de observaciones pasadas (llamado orden de retraso).
- D**: Número de veces que los valores son diferenciadas (grado de diferencia).
- Q**: Tamaño de la ventana de promedio móvil (orden de promedio móvil).

Se construye un modelo de regresión lineal que incluye el número y tipo de términos especificados, y los datos se preparan mediante un grado de diferenciación para hacerlo estacionario, es decir, para eliminar las estructuras estacionales y de tendencia que afectan negativamente a la precisión del modelo.

Se puede usar un valor de 0 en un parámetro para indicar que no se debe usar ese componente del modelo. De esta forma, el modelo ARIMA se puede configurar para realizar la función de un modelo ARMA, e incluso un modelo AR, I o MA simple.

2.6 Modelos ANN

Las ANN (*artificial neural network*) [13-14] son modelos inspirados en las redes neuronales biológicas del cerebro humano. Están constituidas por elementos que se comportan de forma similar a una neurona biológica en sus funciones mas comunes. Las redes neuronales, como profundizaremos más adelante, pueden formarse por una o más capas de neuronas.

Al margen de “parecerse” al cerebro presentan una serie de características propias del cerebro. Por ejemplo las ANN aprenden de la experiencia, generalizan de ejemplos previos y abstraen las características principales de una serie de datos.

Las redes neuronales son capaces de:

- Aprender: adquirir el conocimiento por medio de la experiencia. Se les muestra un conjunto de entradas y ellas mismas se ajustan para producir unas salidas consistentes.
- Generalizar: extender o ampliar un conocimiento. Generalizan automáticamente debido a su propia estructura y naturaleza. De este modo pueden mitigar el efecto que produce en los datos los *outliers* o el ruido.
- Abstraer: son capaces de abstraer la esencia de un conjunto de entradas que aparentemente no presentan aspectos comunes o relativos.

Cada neurona tiene un conjunto de entradas y cada una de estas entradas está ponderada por una serie de pesos. Cada peso se entiende como la importancia del valor de entrada que llega a la neurona.

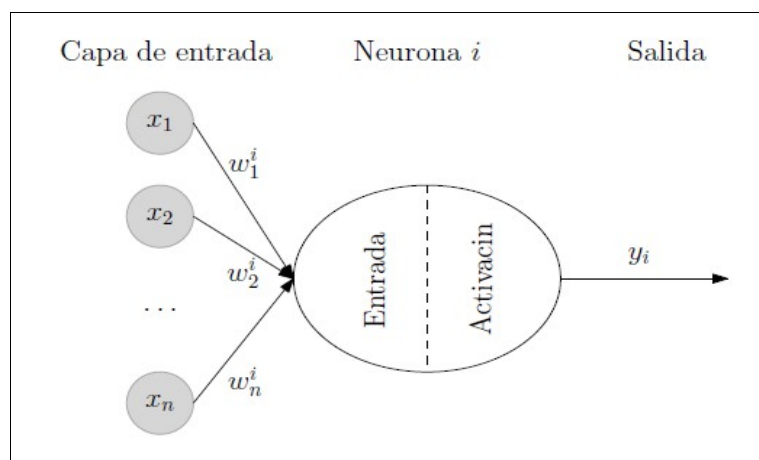


Figura 2.9. Nodo de una red neural

Cada neurona combina los valores de entrada, aplicando sobre ellos una función de entrada. La función de entrada puede ser una función de máximos, mínimos,... aunque la suma ponderada suele ser la función más usada.

$$z(x) = \sum_{j=1}^n x_j w_j^i$$

Figura 2.10. Función de entrada más usada en una red neuronal (suma ponderada)

El valor resultante es procesado por una función de activación, que modula el valor de las entradas para generar el valor de salida de la neurona. La función de activación suelen ser la escalón, lineal, sigmoide o hiperbólica. Destacar que las funciones sigmoide e hiperbólica presentan un comportamiento no lineal.

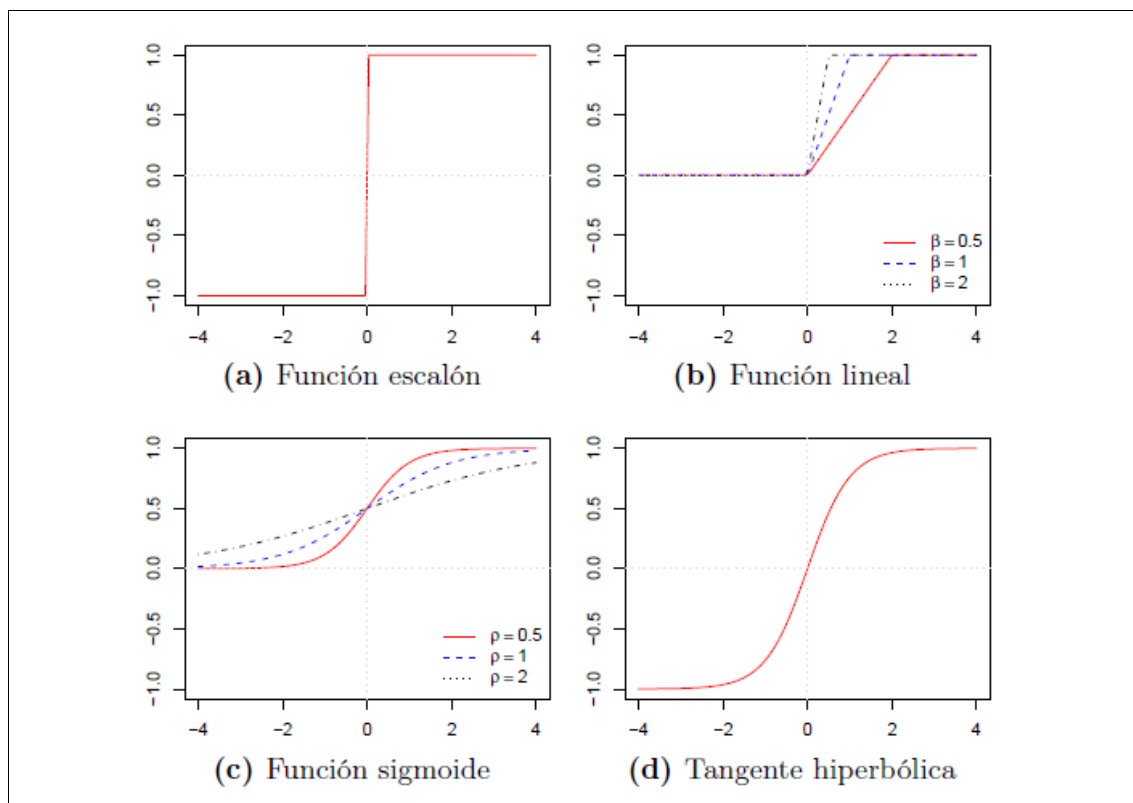


Figura 2.11. Funciones de activación: escalón (a), lineal (b), sigmoide (c) e hiperbólica (d).

El valor de salida de la función de activación, generalmente se propaga a la siguiente capa de conexiones neuronales o puede ser el valor definitivo resultante de la red neuronal.

Como ya hemos dicho antes, las redes neuronales pueden formarse por más de una capa. Cada capa puede tener su propia configuración en cuanto a pesos, función de entrada y función de activación. Podemos encontrar redes con más de una capa o más de una salida. Cuantas más capas mejor predicción obtendremos pero necesitaremos también más coste computacional.

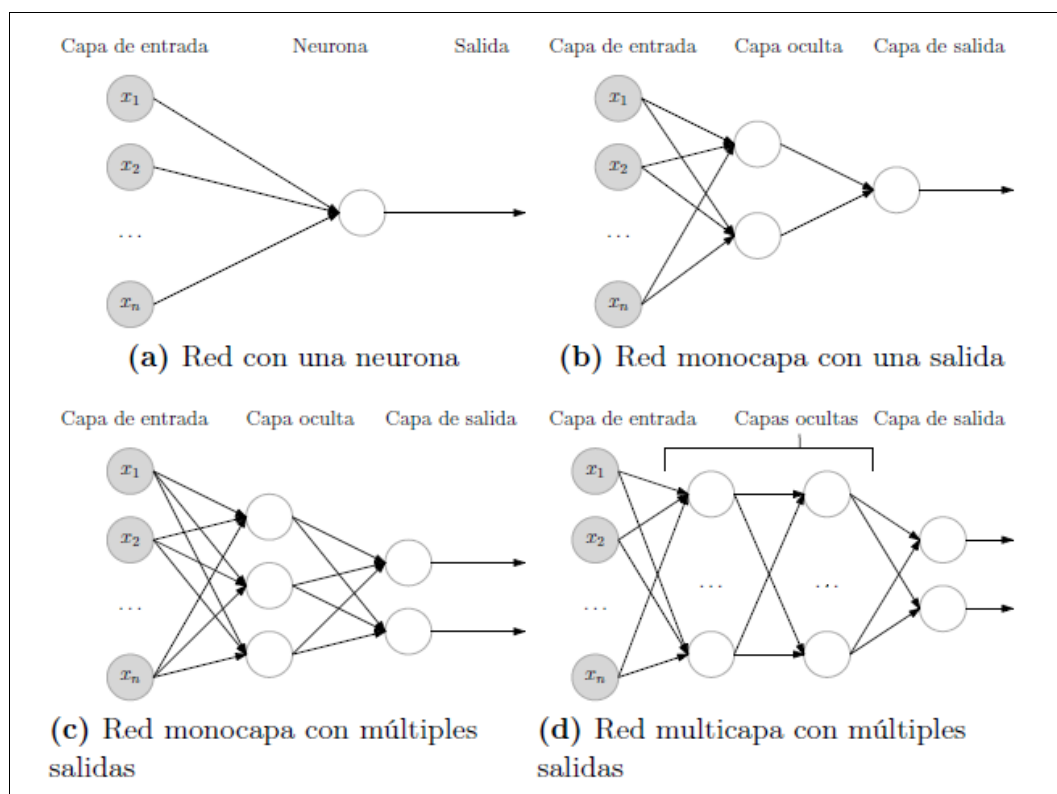


Figura 2.12. Diferentes arquitecturas de redes neuronales FNN.

Estos ejemplos de la figura anterior (2.12) muestran distintas arquitecturas de redes neuronales del tipo FNN (*feedforward neural network*). Este tipo de redes no producen bucles, la información siempre avanza y no se retroalimenta. De aquí destaca la CNN (*convolutional neural network*) usada sobretodo para clasificación de imágenes.

Pero hay otro tipo de arquitecturas con bucles retroalimentados, donde salidas de capas neuronales se reutilizan como entrada del sistema. Son las llamadas RNN (*recurrent neural network*) y son capaces de tener conocimiento de los datos en base al tiempo y el orden. Entre ellas destaca la LSTM (*long short-term memory*), usadas en reconocimiento de texto y predicción de series temporales, como es en nuestro caso de estudio.

Pero no todo son bondades en este tipo de modelos automáticos, cuando entrenamos nuestros modelos intentamos obtener una salida conocida a una entrada de valores conocidos. Así que estos modelos pueden tender a sobreentrenarse, aprender los casos particulares que han adquirido en el entrenamiento y ser incapaces de reconocer nuevos datos de entrada. Este problema se llama también *overfitting*.

2.6.1 Red neuronal CNN

Las redes neuronales del tipo CNN (*convolution neural network*) son una extensión de las FNN (*feedforward neural network*), muy eficaces para procesar datos visuales y otros datos bidimensionales, son muy similares a las redes neuronales ordinarias multicapa, lo que hace que tengan un entrenamiento más corto al no ser retroalimentadas. Cada uno de los nodos de una capa no están conectados con las anteriores, sino que cada capa funciona como un filtro. Los filtros trabajan como una ventana que se desliza por la imagen en busca de patrones.

Las redes neuronales CNN se forman de las siguientes capas o etapas [15]:

- Capa de convolución: un filtro o ventana llamada *kernel* que se desliza sobre la imagen, visualizando unos pocos *píxeles* a la vez (por ejemplo, 3x3 o 5x5). La operación de convolución es un producto de los valores de *píxeles* originales con pesos definidos en el filtro. Los resultados se resumen en un número que representa todos los *píxeles* que observó el filtro.
- Capa de activación: la capa de convolución genera una matriz que es mucho más pequeña en tamaño que la imagen original. Esta matriz se ejecuta a través de una capa de activación, normalmente con una función de activación ReLu.
- Capa de Pooling: es el proceso de disminución de muestras y reducción del tamaño de la matriz. Se pasa un filtro sobre los resultados de la capa anterior y selecciona un número de cada grupo de valores. Esto permite que la red entrene mucho más rápido, enfocándose en la información más importante en cada característica de la imagen.
- Capa Fully connected: una estructura multicapa tradicional. Su entrada es un vector unidimensional que representa la salida de las capas anteriores. Su salida es una lista de probabilidades para diferentes etiquetas posibles adjuntas a la imagen. La etiqueta que recibe la mayor probabilidad es la decisión de clasificación.

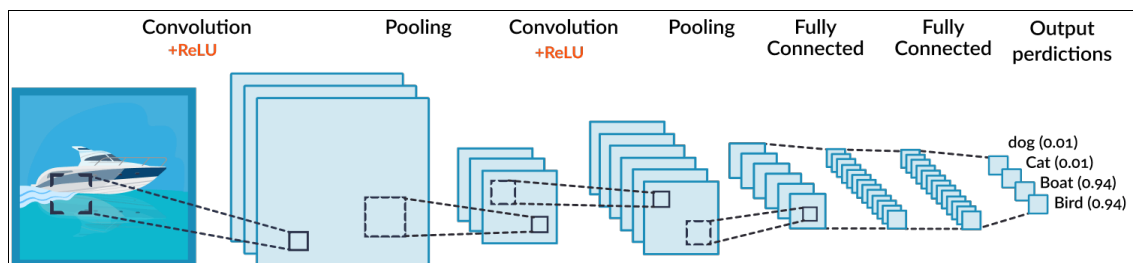


Figura 2.13. Capas de una red neuronal del tipo convolucional (CNN)

Aunque su función más conocida es la de clasificar imágenes, también son útiles para la predicción de series temporales. Las imágenes no son más que matrices de *píxeles*, así que si la red neuronal CNN es capaz de trabajar con imágenes, si transformamos la matriz 2D a un vector 1D, la red también será capaz de trabajar con un vector que represente una serie temporal [16]. Así que los filtros que trabajan como ventanas deslizantes en busca de patrones en una imagen, también pueden buscar patrones en nuestra serie temporal.

2.6.2 Red neuronal LSTM

Las redes neuronales del tipo LSTM (*long-short term memory*) son una extensión de las RNN (*recurrent neural network*), que básicamente amplían su memoria para aprender de experiencias importantes que han pasado a lo largo del tiempo. Las LSTM permiten a las RNN recordar sus entradas durante un período de tiempo. Esto se debe a que LSTM contiene información en una memoria.

Esta memoria se puede ver como una “celda” bloqueada, donde “bloqueada” significa que la neurona decide si almacenar o eliminar información dentro, en función de la importancia que asigna a la información que está recibiendo. La asignación de importancia se decide a través de los pesos, que también se aprenden mediante el algoritmo. Esto lo podemos ver como que aprende con el tiempo qué información es importante y cuál no.

Las celdas LSTM tiene tres puertas o *gates* que dan acceso a esta información: puerta de entrada (*input gate*), puerta de olvido (*forget gate*) y puerta de salida (*output gate*). Estas puertas determinan si se permite o no una nueva entrada, se elimina la información porque ya no es importante o se deja que afecte a la salida en el paso de tiempo actual.

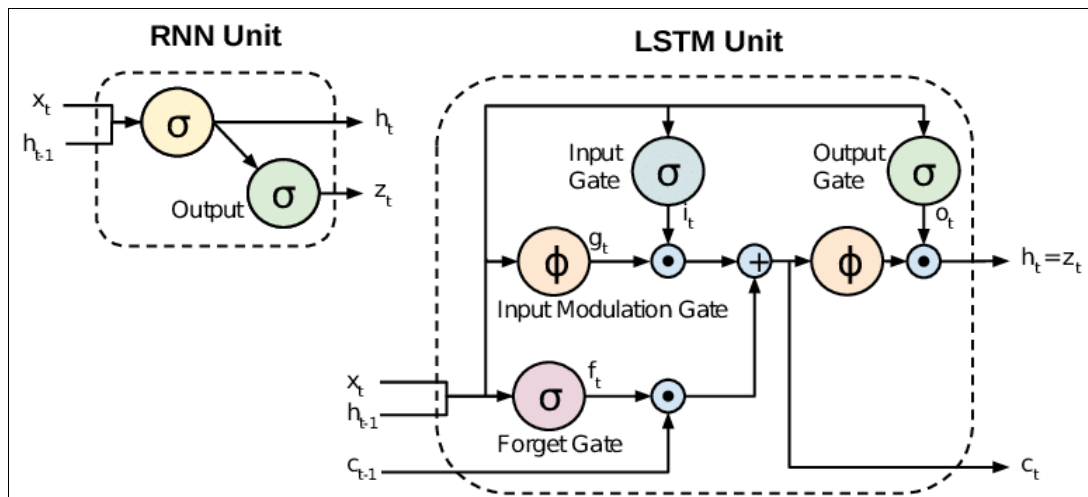


Figura 2.14. Esquema de una celda RNN y una LSTM

Las puertas en una LSTM son análogas a una forma sigmoide, lo que significa que van de 0 a 1. El hecho de que sean análogas a una función de activación sigmoide, permite incorporarlas al proceso de retroalimentación.

Existen otras implementaciones RNN parecidas como las GRU (*gated recurrent unit*). Las capas GRU aparecieron en el 2014 y usan el mismo principio que LSTM, pero están simplificadas de manera que tienen menos puertas o *gates*, su rendimiento esparecido al de las LSTM pero son más eficientes en cuanto a coste computacional.

5. Bibliografía

- [1] Noticia: “Almacenamiento de energía eléctrica a gran escala” (27-feb-2020)
<http://www.agenex.net/es/documentacion/381-almacenamiento-de-energia-electrica-a-gran-escala.html>
- [2] Noticia: “Empresas españolas logran almacenar a gran escala electricidad de fuentes renovables” (27-feb-2020)
<https://www.energias-renovables.com/panorama/empresas-espanolas-logran-almacenar-a-gran-escala-20140605>
- [3] Web: “Demanda de energía eléctrica en tiempo real” (27-feb-2020)
<https://demanda.ree.es/visiona/peninsula/demanda/total>
- [4] Web: “Previsión meteorológica y clima Madrid” (27-feb-2020)
<https://www.weather-es.com/es/espana/madrid-el-tiempo-en-enero>
- [5] Noticia: “Smartcity Málaga, 10 años después” (29-feb-2020)
<https://www.lavanguardia.com/natural/si-existe/20200128/473098862713/malaga-smart-city-endesa-laboratorio-living-lab-diez-anos-brl.html>
- [6] Libro: A. Palit, D. Popovic. “Computational Intelligence in Time Series Forecasting” (2005). (págs: 17-24)
- [7] Web: “Introduction to stationarity”
<https://www.analyticsvidhya.com/blog/2018/09/non-stationary-time-series-python>
- [8] Web: “Decomposition of time series” (wikipedia)
https://en.wikipedia.org/wiki/Decomposition_of_time_series
- [9] Libro: R.J Hyndman, G. Athanasopoulos. “Forecasting: Principles and Practice” (2008). Capítulo 6: “Time series decomposition”
- [10] Libro: Chrios Chatfield. “Time-Series Forecasting” (2000). Capítulo 6: “A Comparative Assessment of Forecasting Methods”
- [11] Web: “Understanding deep learning models for time series forecasting and their advantages over traditional models like ARIMA” (20-mar-2020)
<https://towardsdatascience.com/deep-learning-for-time-series-and-why-deep-learning-a6120b147d60>
- [12] Web: “How to Create an ARIMA Model for Time Series Forecasting” (20-mar-2020)
<https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>
- [13] Libro: J. Gironés, J. Casas, J. Minguillón, R. Caihuelas. “Minería de datos: modelos y algoritmos” (UOC, 2017). Capítulo 12: “Redes neuronales”
- [14] Web: “Redes neuronales ¿qué son?” (21-mar-2020)
<https://medium.com/@williamkhepri/redes-neuronales-que-son-a64d022298e0>
- [15] Web: “Convolutional Neural Network Tutorial: From Basic to Advanced” (21-mar-2020)
<https://missinglink.ai/guides/convolutional-neural-networks/convolutional-neural-network-tutorial-basic-advanced/>

[16] Vídeo: J. Ordóñez. “Deep Learning para el análisis de series temporales” (21-mar-2020)
<https://www.youtube.com/watch?v=7gslkXpZx9E>