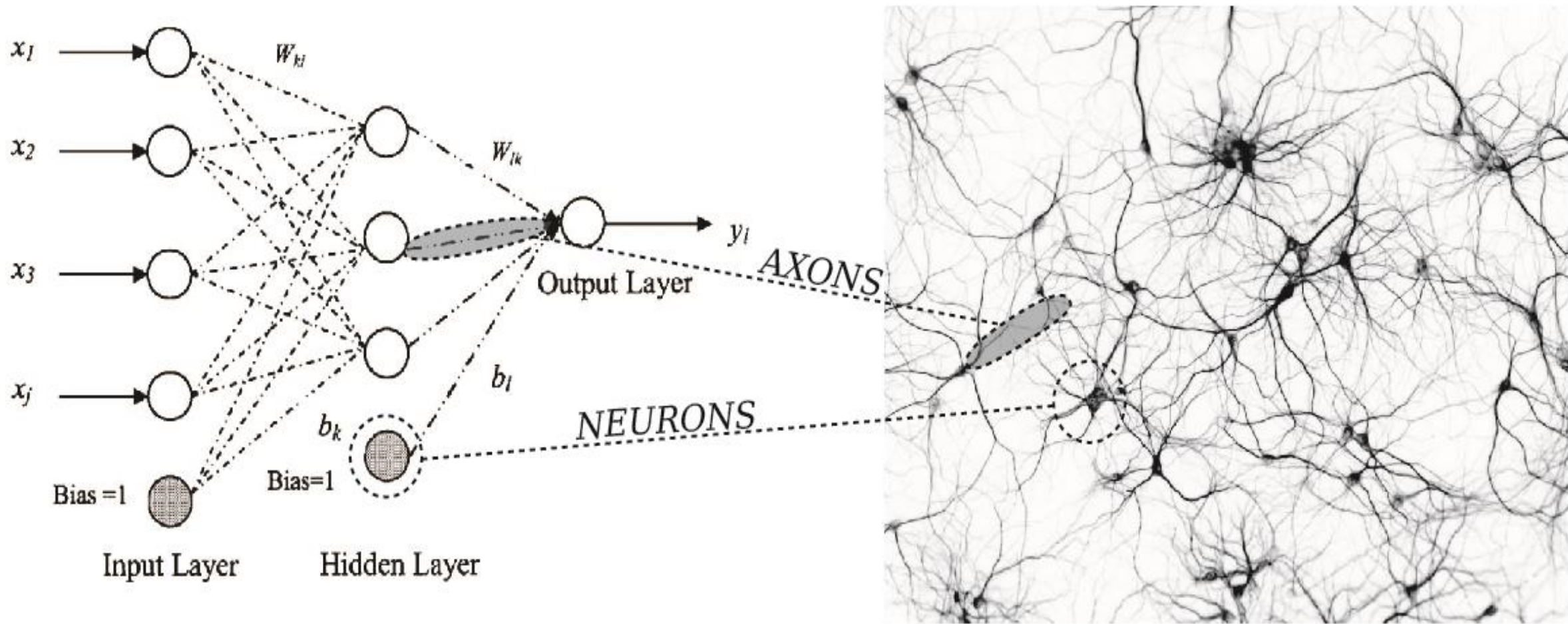
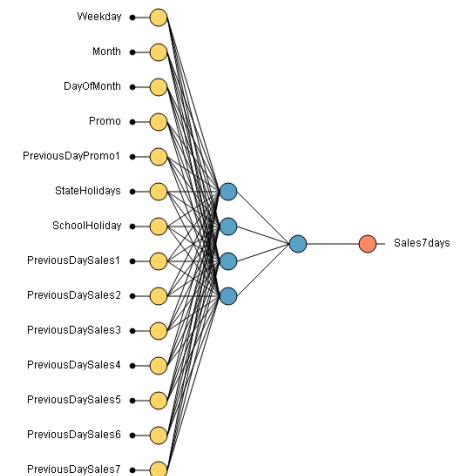
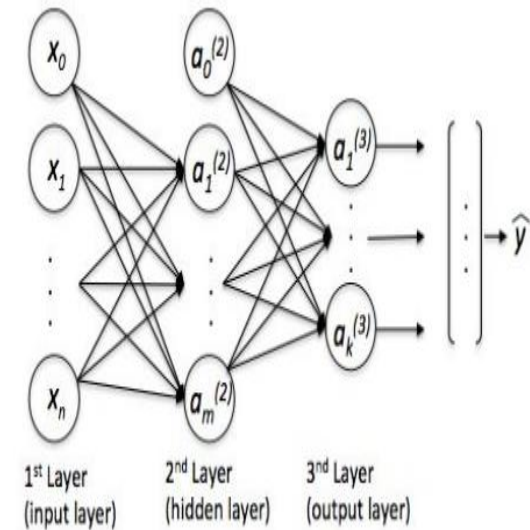


Redes Neuronales en las series temporales



Preámbulo

- Las redes neuronales contrastan respecto a los métodos vistos anteriormente: es una técnica no paramétrica. Si bien podría llegar a ser la técnica que brinde los mejores resultados, muchas veces los procesos de estimación suelen ser mucho más complejos que todos los anteriores... también no necesariamente se llegará a obtener los mejores resultados... el tamaño de la serie T influye mucho en este punto.
- El presente capítulo explicará los fundamentos de una red neuronal, su forma funcional en el ámbito transversal, su interpretación, para finalmente hacer el enlace con la aplicación de un red neuronal en las series temporales.
- Aunque yo mismo lo menciono como “modelos de la 3era generación”, además de las redes, estudiaremos los SVM como otra forma de investigar la estructura oculta de los datos.



Preámbulo

- Si nos preguntamos, ¿una red neuronal es un análisis de Inteligencia Artificial?, entonces, ¿la podríamos aplicar como un método de Estadística...o no?



Preámbulo

En este capítulo se intentará dar una diferenciación en lo que se conoce como “Estadística”, “Machine Learning”, “Deep Learning” e “Inteligencia Artificial”. No entraremos a contrastar también lo que se entiende por Minería de datos...



Índice

1

Introducción

4

Neurona biológica vs
neurona artificial

2

Estadística, Machine
Learning, Deep Learning e
Inteligencia Artificial

5

Redes Neuronales
Artificiales

3

¿Por qué utilizar RNA en
series de tiempo?

6

Aplicación de las redes en
las series de tiempo

Índice

1

Introducción

Introducción

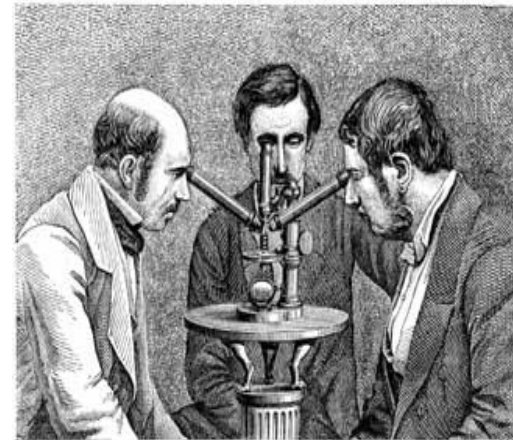
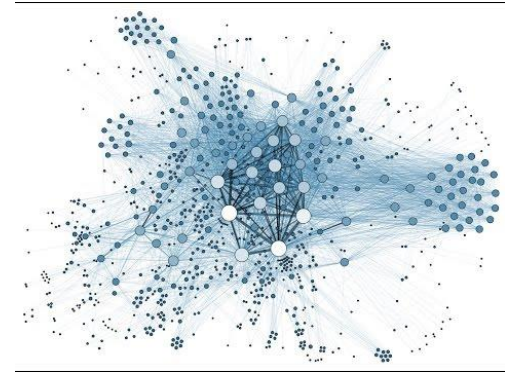
Dado que muchas de las series de tiempo que son de interés de análisis tienen una naturaleza no lineal, se hace necesario utilizar otras técnicas fuera de las clásicas para realizar la predicción de estas y así obtener modelos más eficientes.

En los últimos años las Redes Neuronales Artificiales (*RNA* o *NNA* en inglés) han sido exitosamente aplicadas como herramienta en la predicción de series de tiempo en un amplio rango de problemas en áreas de comercio, economía, industria, aeronáutica, ciencias, etc...

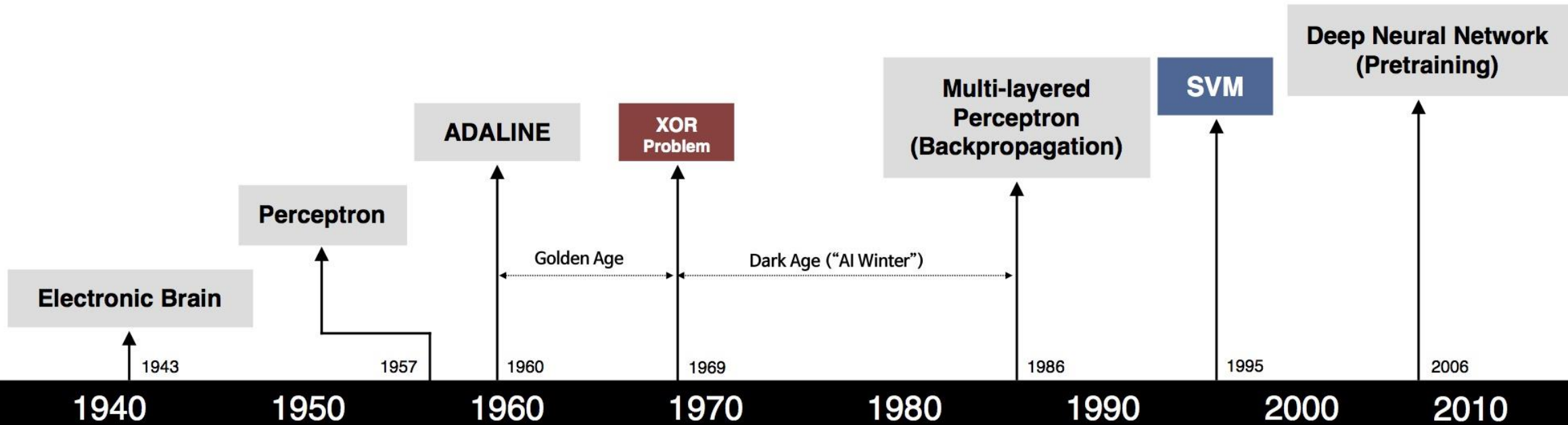
Las RNA son una rama de la Inteligencia Artificial que consiste en el aprendizaje y procesamiento automático, inspirado en la forma en que funciona el sistema nervioso biológico. Veamos un poco de historia:

https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html

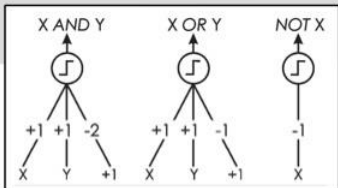
<https://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/>



Introducción



S. McCulloch – W. Pitts



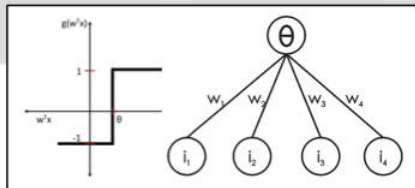
- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



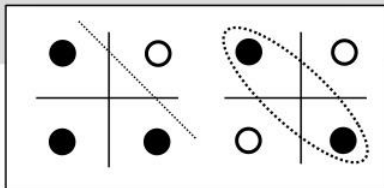
B. Widrow – M. Hoff



- Learnable Weights and Threshold



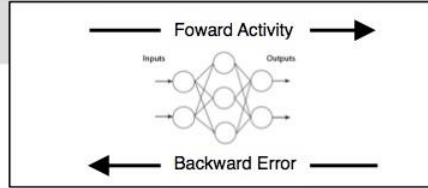
M. Minsky – S. Papert



- XOR Problem



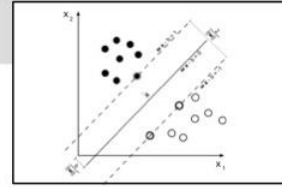
D. Rumelhart – G. Hinton – R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



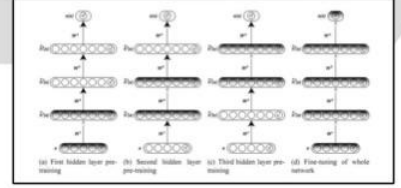
V. Vapnik – C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention



G. Hinton – S. Ruslan



- Hierarchical feature Learning

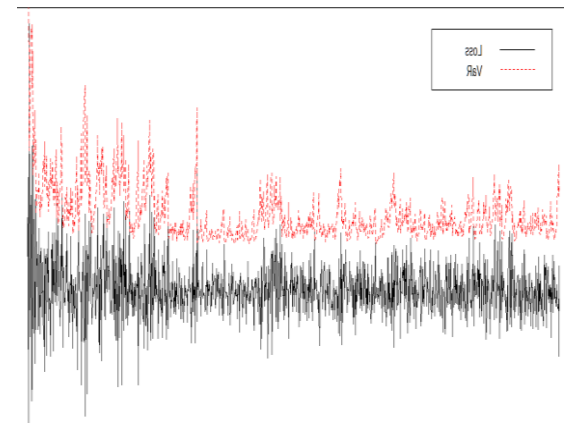
Introducción

El objetivo de la utilización de RNA es conseguir respuestas similares a las que es capaz de dar el cerebro, que se caracterizan por su generalización y robustez. Las RNA son útiles en la predicción de series de tiempo dado que a diferencia de los métodos econométricos clásicos, son capaces de capturar las relaciones lineales y no lineales entre los datos debido a su estructura no lineal que permite un modelo con más grados de libertad.



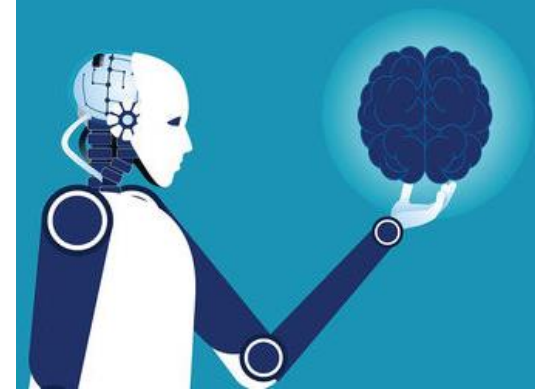
Las redes neuronales han sido utilizadas exitosamente en muchos tipos de problemas de predicción debido a que son capaces de modelar y predecir series de tiempo, lineales y no lineales, con un alto grado de precisión, además de capturar cualquier tipo de interrelación entre los datos y no requerir conocimiento previo respecto del problema que se está modelando.

La gran debilidad es que necesitan de un mayor tamaño de datos para poder lograr mejores resultados...

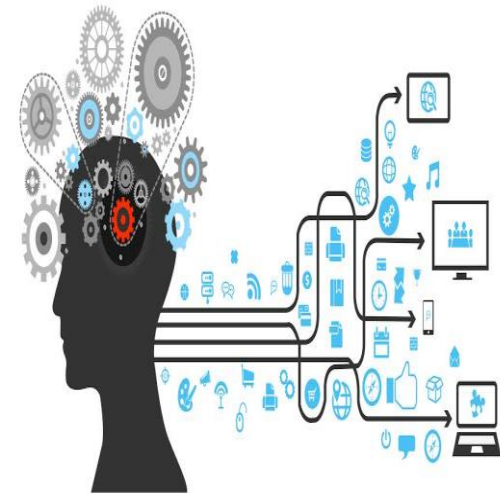


Introducción

La Inteligencia Artificial (IA) se ha establecido como un área de la ciencia del conocimiento que consiste en el aprendizaje y procesamiento automático de la información, inspirado en la forma en que funciona el sistema nervioso biológico. Las Redes Neuronales Artificiales (RNA) constituyen una rama importante dentro de la IA, y tienen como objetivo resolver un problema determinado basado en la simulación de un sistema neuronal biológico simplificado. Es decir, el conocimiento es adquirido mediante un método de aprendizaje a partir de datos representativos del problema en cuestión y se genera una solución a este.

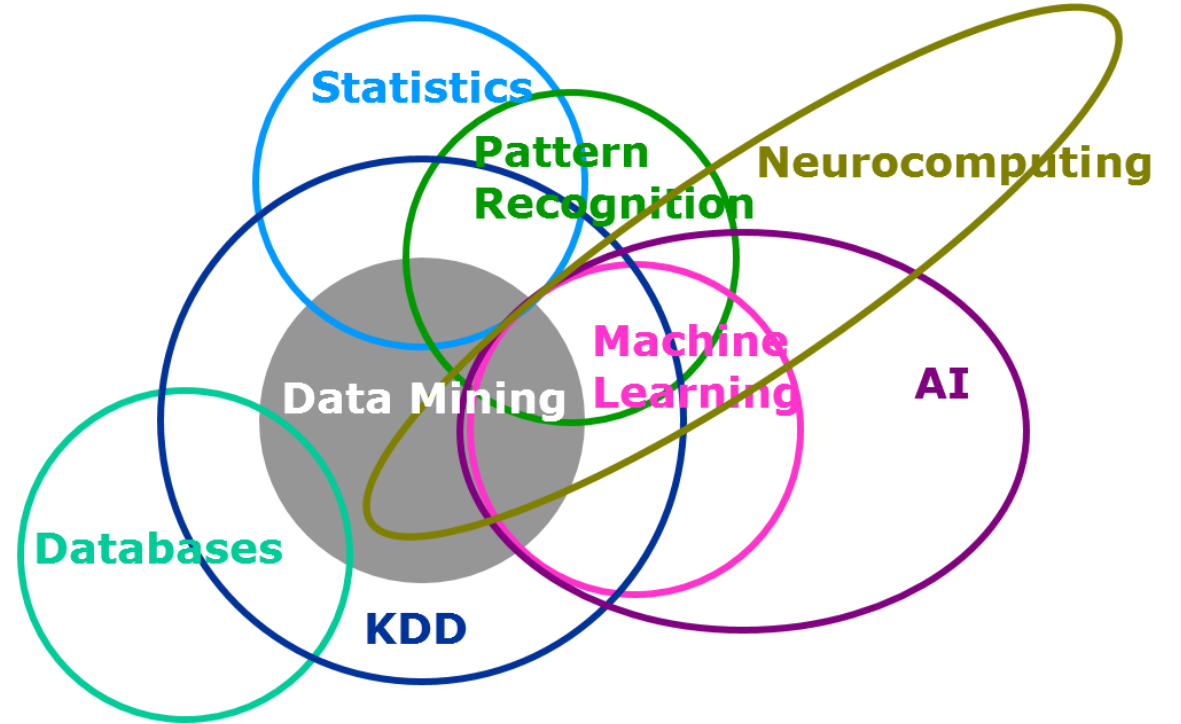
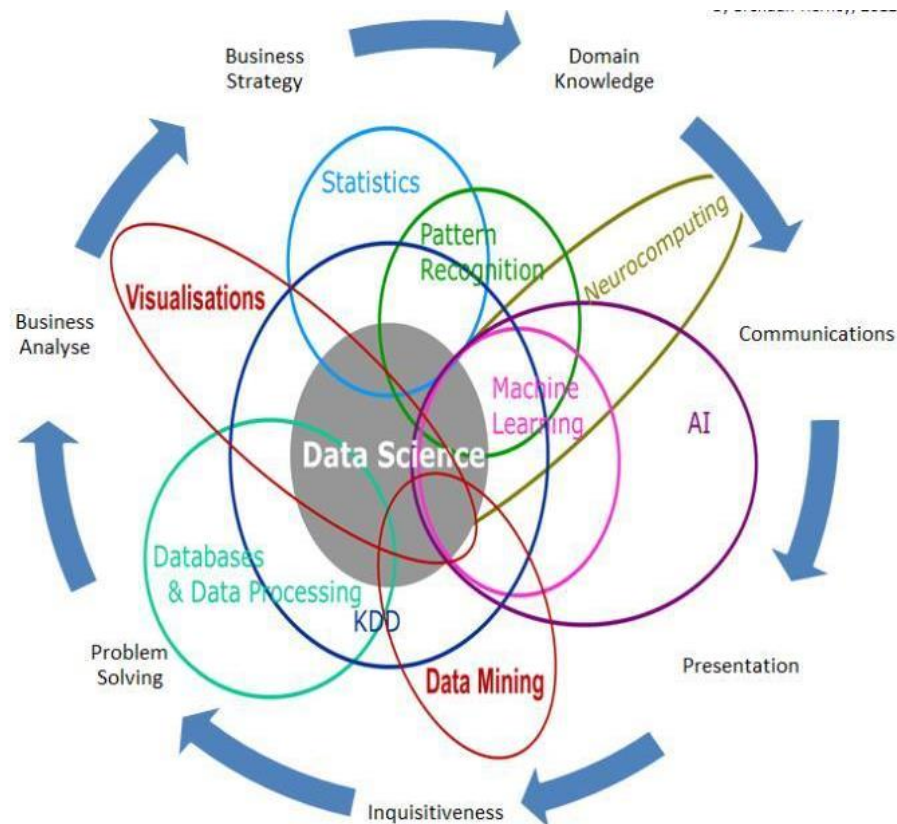


Los modelos de RNA no requieren de un especialista en el problema para llegar a la solución debido a que los métodos de aprendizaje utilizados son capaces de “aprender de la experiencia”, reconociendo de buena forma los patrones existentes en los datos. Por lo tanto, no es necesario conocer las reglas que determinan la solución del problema, sino que mediante un acercamiento de **repetición**, el modelo neuronal ajusta cada uno de sus parámetros de manera de ofrecer un modelo que sea capaz de reconocer los patrones presentados, tolerar errores y generalizar ante datos que no fueron presentados.



Introducción

Antes de seguir con la aplicación de las RNA en las series temporales, es importante discutir lo que entendemos por Estadística, Machine Learning, Deep Learning e Inteligencia Artificial...



Índice

1

Introducción

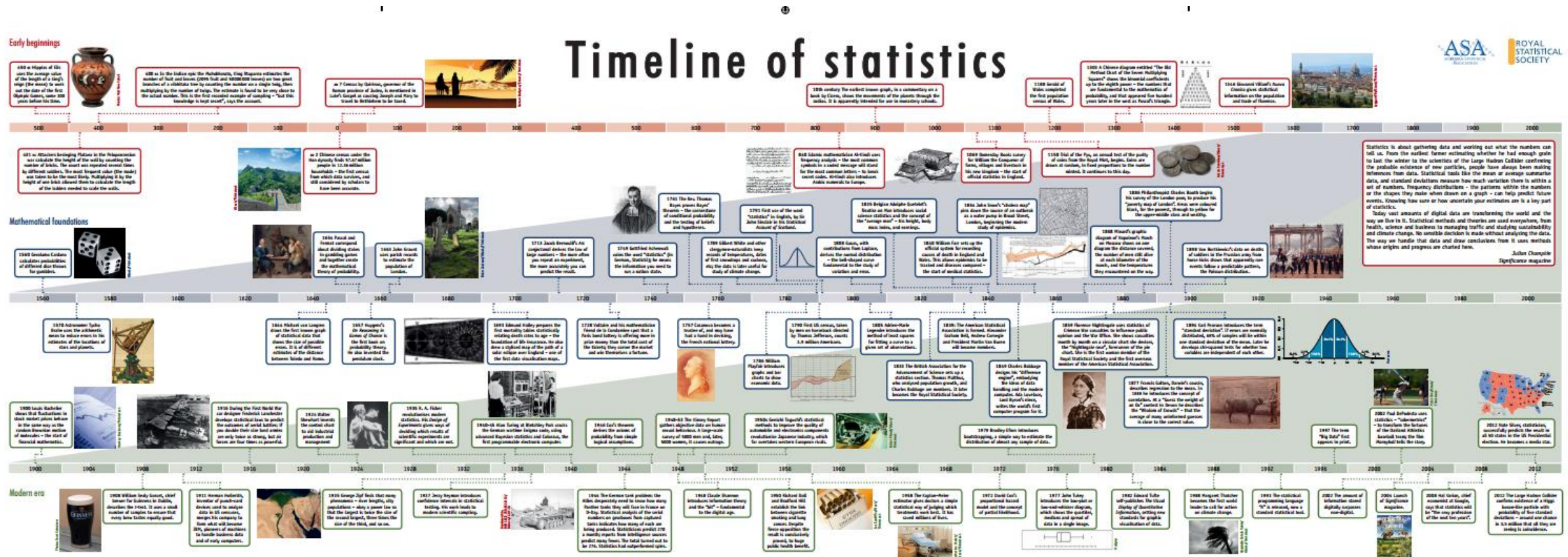
2

Estadística, Machine Learning, Deep Learning e Inteligencia Artificial

Estadística, Machine & Deep Learning e Inteligencia Artificial

Empecemos por lo básico: ¿qué es Estadística?

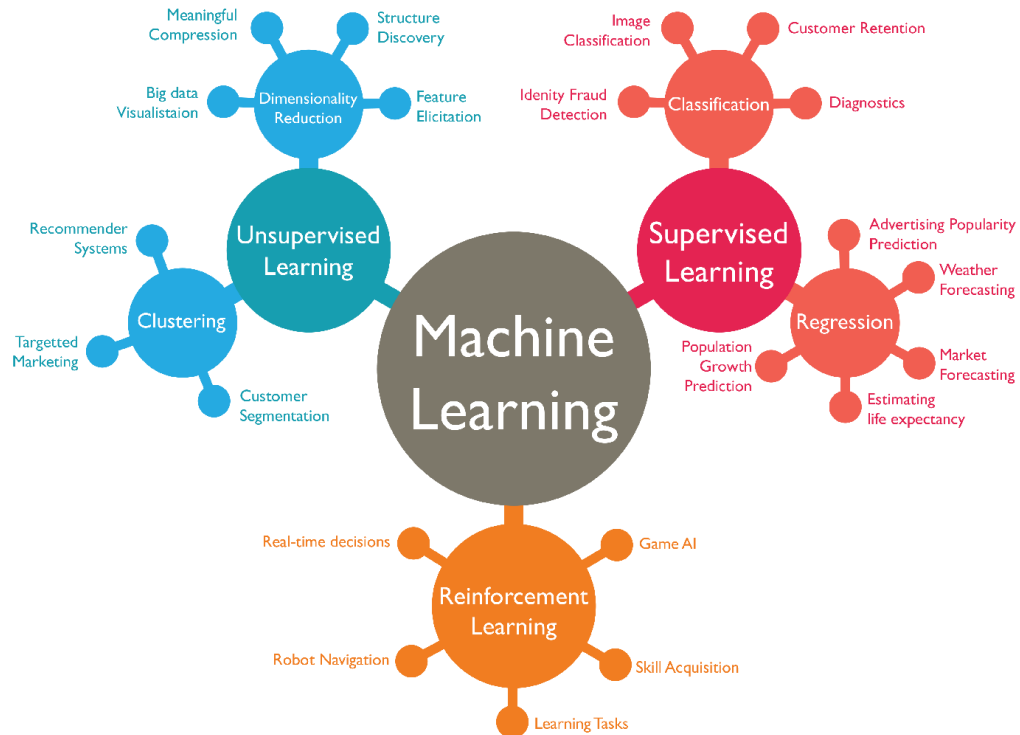
R/ La estadística es la ciencia del **error**, y todo fue creado para explicar el error como tal (lo que se observa y lo que se espera). Este curso es meramente la aplicación de la Estadística, en donde su optimiza o se estima algo en un momento dado para minimizar el error u otro criterio.



Estadística, Machine & Deep Learning e Inteligencia Artificial

Ahora, ¿qué es el “Machine Learning”?

R/ El Machine Learning es una aplicación de *inteligencia Artificial* (IA) que proporciona a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia sin ser programado explícitamente. El aprendizaje automático se centra en el desarrollo de programas informáticos que pueden acceder a los datos y utilizarlos, aprendiendo por ellos mismos...



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset machine learning – have created ever larger disruptions.

Estadística, Machine & Deep Learning e Inteligencia Artificial

¿Diferencia entre Estadística y Machine Learning?

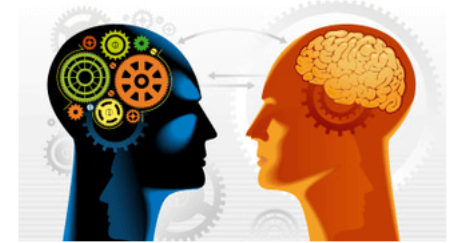
El Machine Learning es un algoritmo que puede aprender de los datos sin depender de la programación, se basada en reglas informáticas. El modelado estadístico es una formalización de las relaciones entre variables en los datos en forma de ecuaciones matemáticas. El aprendizaje automático tiene que ver con predicciones, aprendizaje supervisado, aprendizaje no supervisado, etc. Las estadísticas son sobre muestra, población, hipótesis, etc...

Sin embargo, ambos tienen el mismo :

“They are both concerned with the same question: how do we learn from data?”

- “Machine learning is essentially a form of applied statistics”
- “Machine learning is glorified statistics”
- “Machine learning is statistics scaled up to big data”
- “The short answer is that there is no difference”

Machine Learning Vs. Statistics

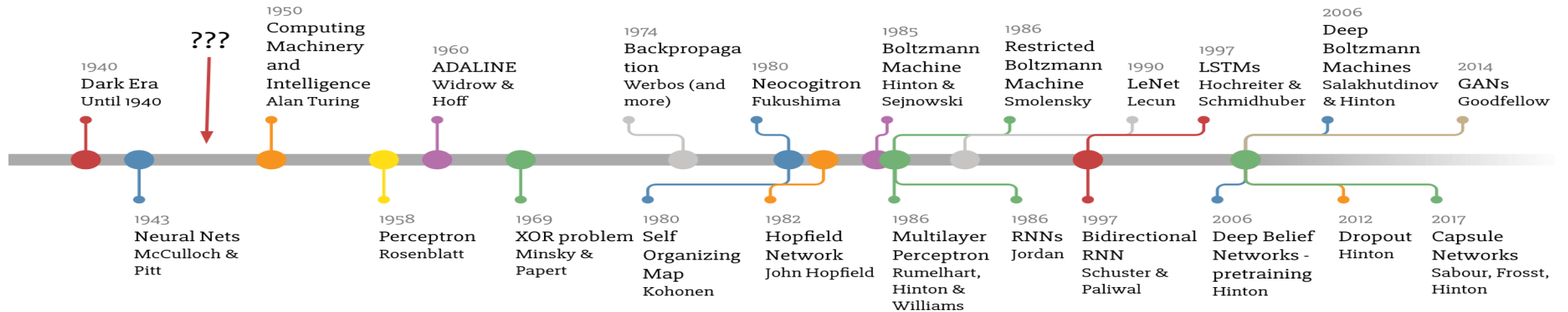


Estadística, Machine & Deep Learning e Inteligencia Artificial

¿Qué es el Deep Learning?

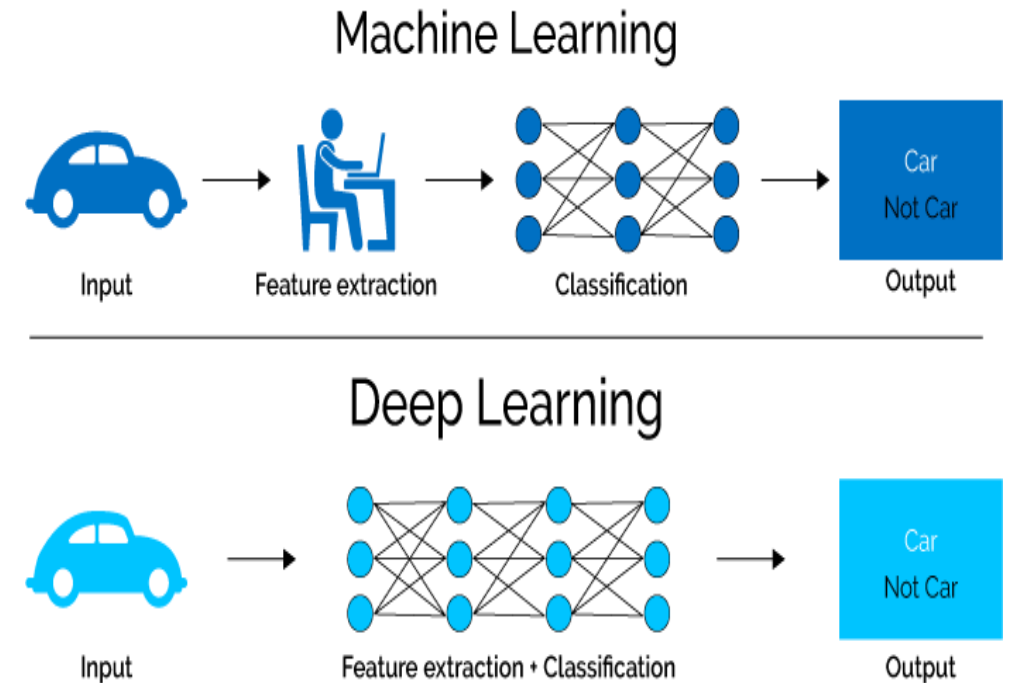
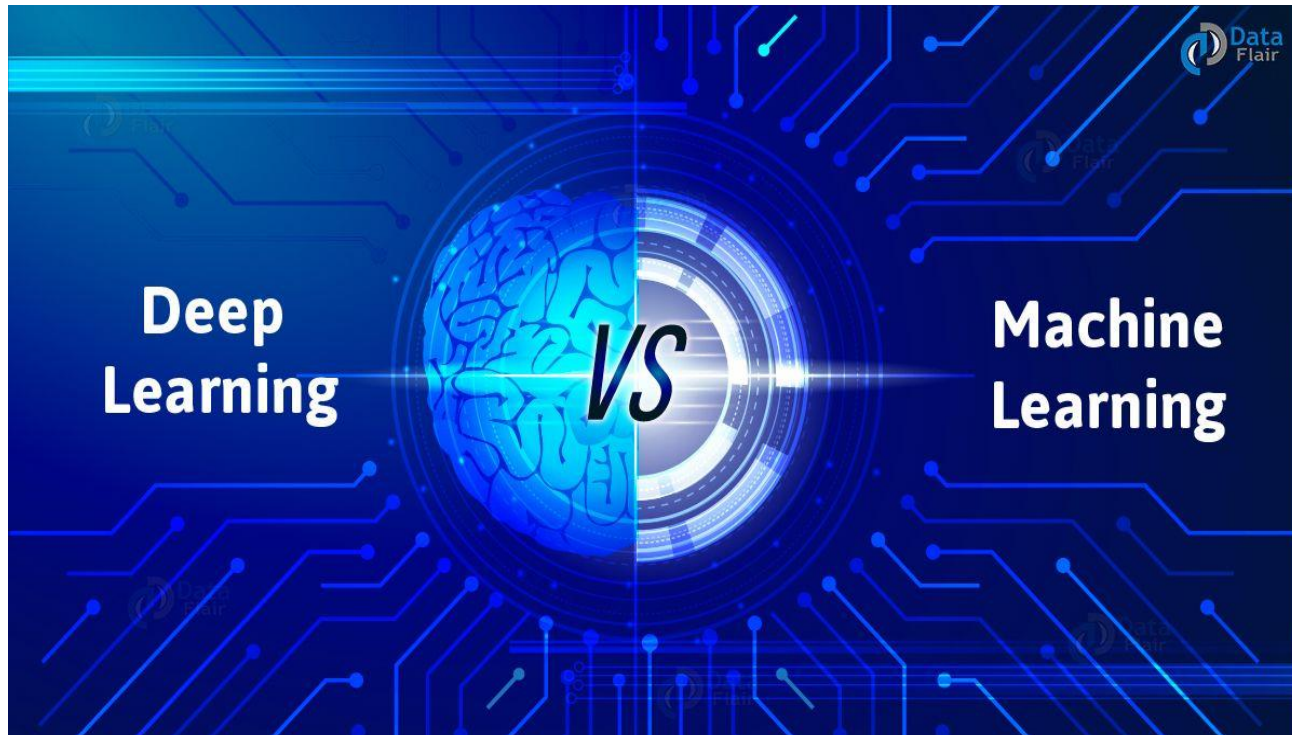
R / Deep Learning is an artificial intelligence function that imitates the workings of the human brain in processing data and creating patterns for use in decision making. Deep learning is a subset of machine learning in Artificial Intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled.

Deep Learning Timeline



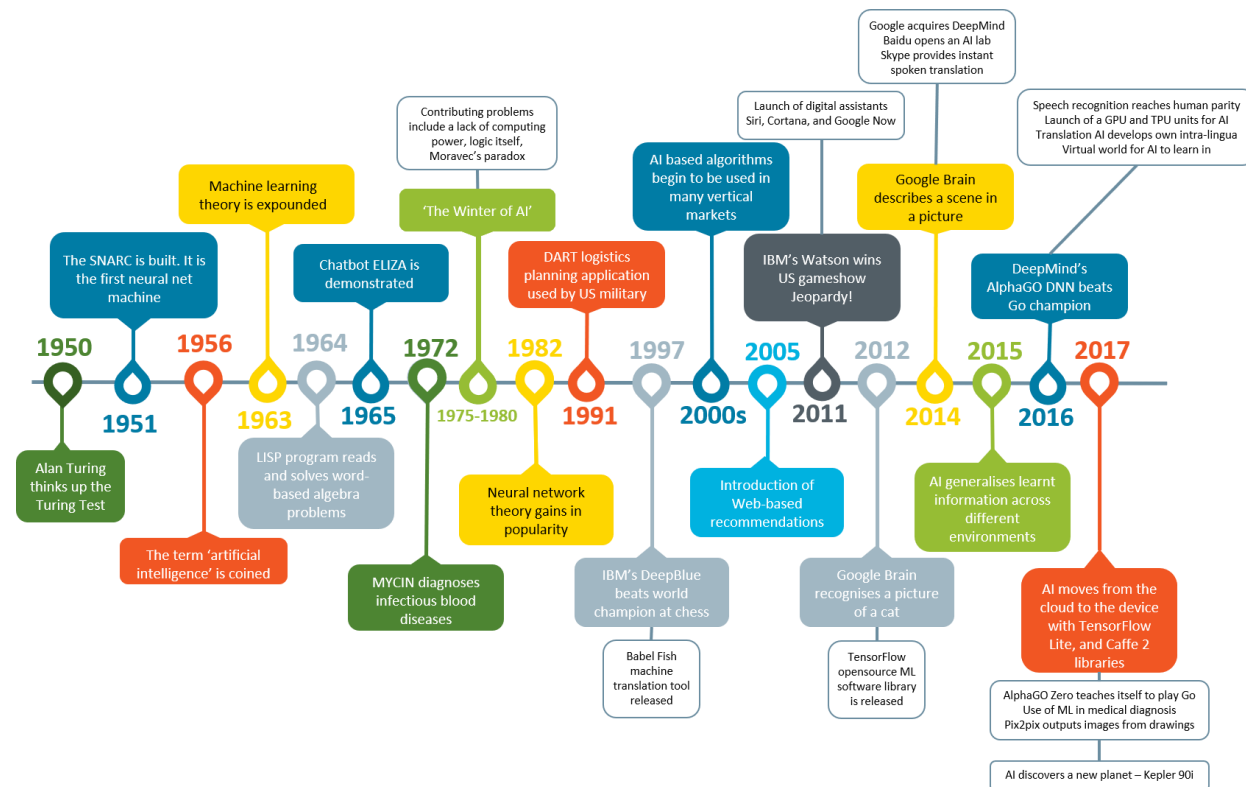
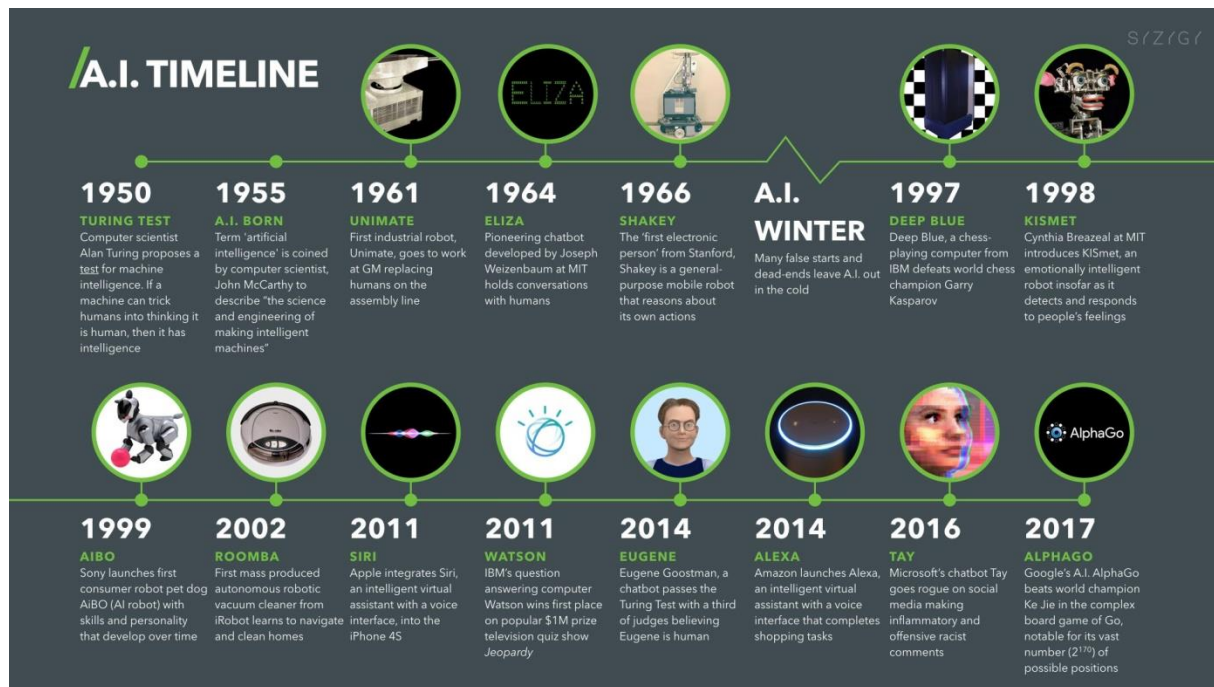
Estadística, Machine & Deep Learning e Inteligencia Artificial

¿Diferencias entre el Machine y el Deep learning?



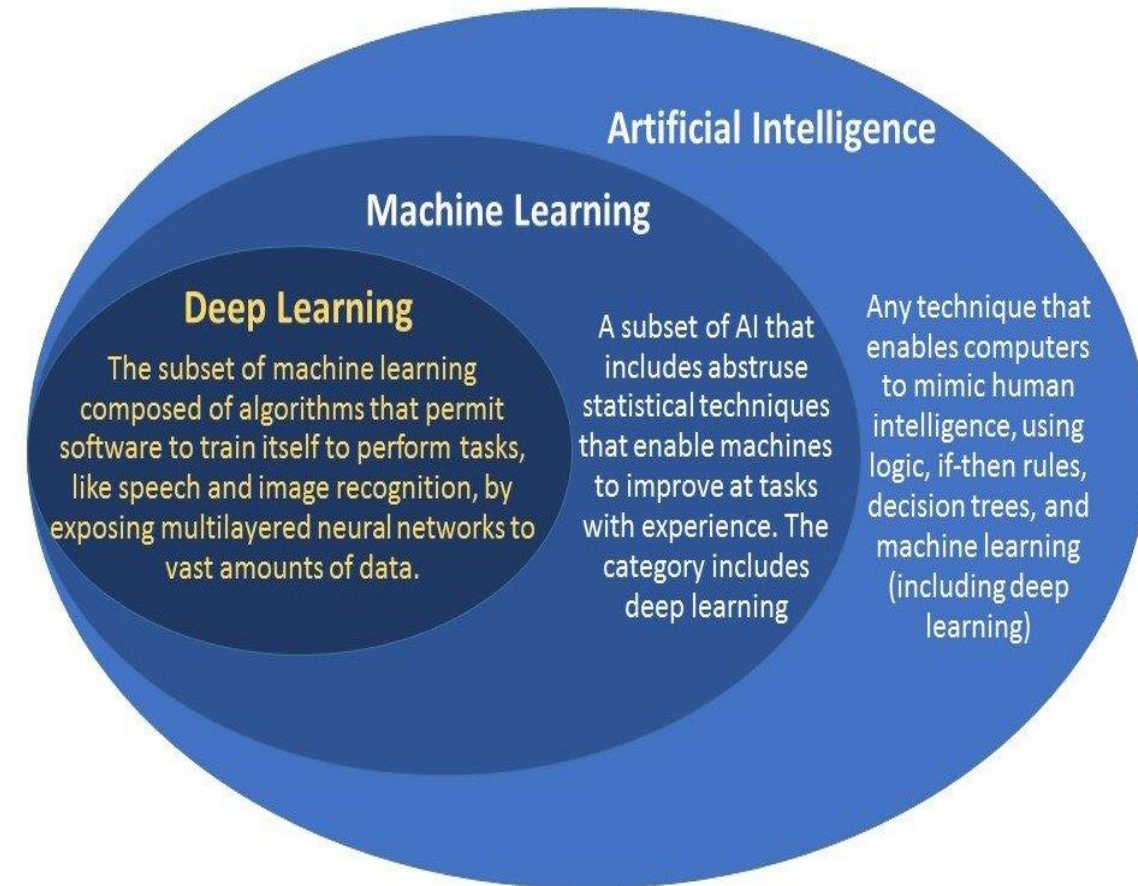
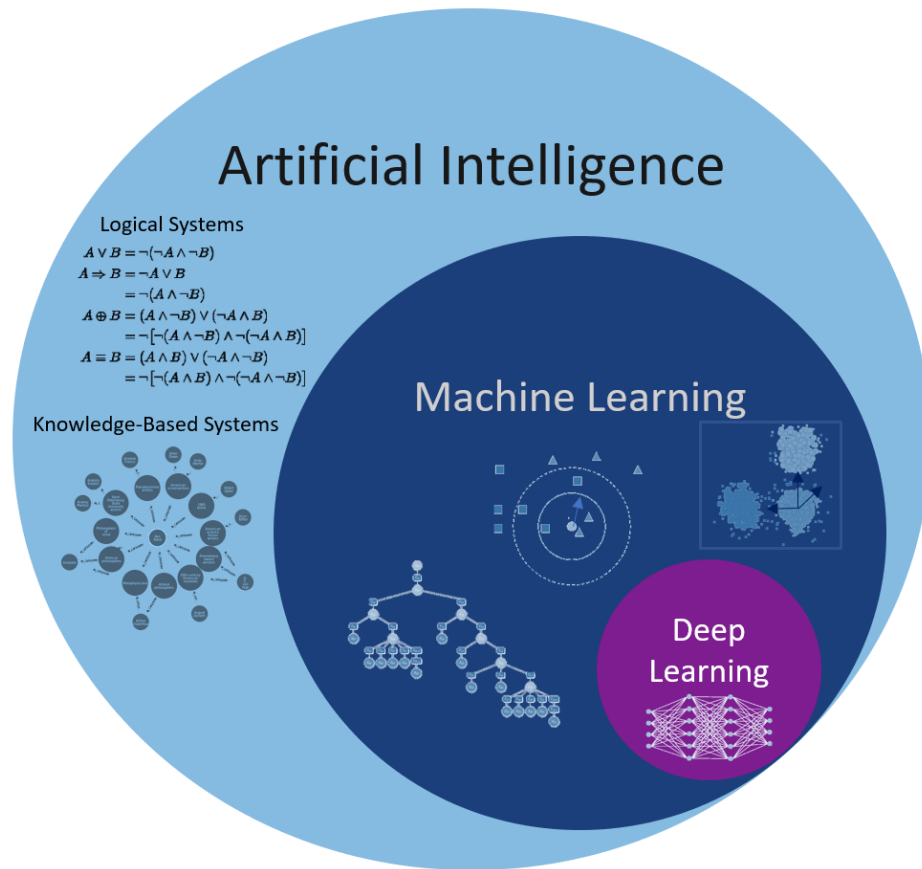
Estadística, Machine & Deep Learning e Inteligencia Artificial

¿Qué es el análisis por Inteligencia Artificial? R / AI (inteligencia artificial) es la simulación de los procesos de inteligencia humana mediante máquinas, especialmente sistemas informáticos. Estos procesos incluyen el aprendizaje (la adquisición de información y las reglas para usar la información), el razonamiento (el uso de reglas para llegar a conclusiones aproximadas o definidas) y la autocorrección. Las aplicaciones particulares de la IA incluyen sistemas expertos, reconocimiento de voz y visión artificial.



Estadística, Machine & Deep Learning e Inteligencia Artificial

- Conclusión del Machine Learning, Deep Learning y AI.



Estadística, Machine & Deep Learning e Inteligencia Artificial

ARTIFICIAL INTELLIGENCE

IS NOT NEW

ARTIFICIAL INTELLIGENCE

Any technique which enables computers to mimic human behavior



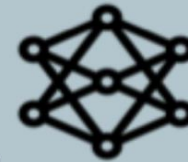
MACHINE LEARNING

AI techniques that give computers the ability to learn without being explicitly programmed to do so



DEEP LEARNING

A subset of ML which make the computation of multi-layer neural networks feasible



1950's

1960's

1970's

1980's

1990's

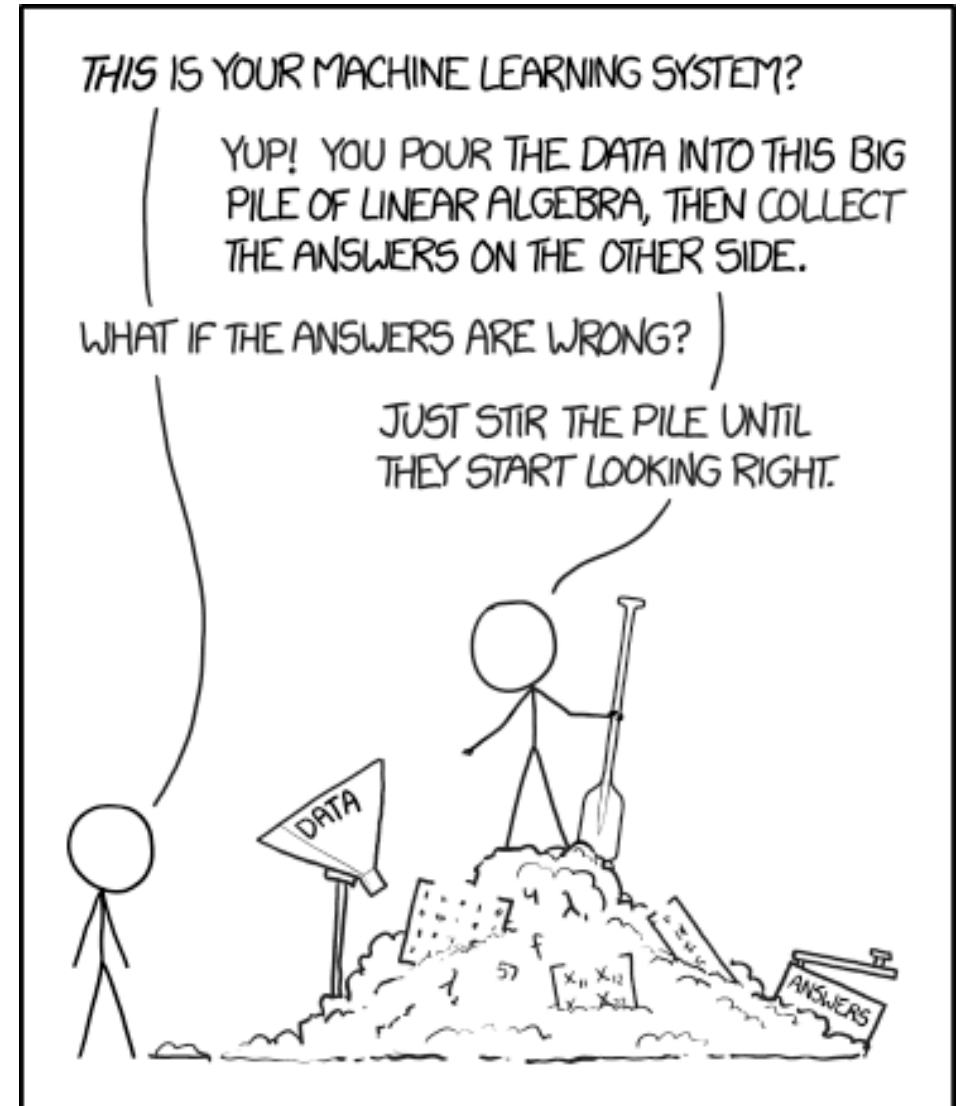
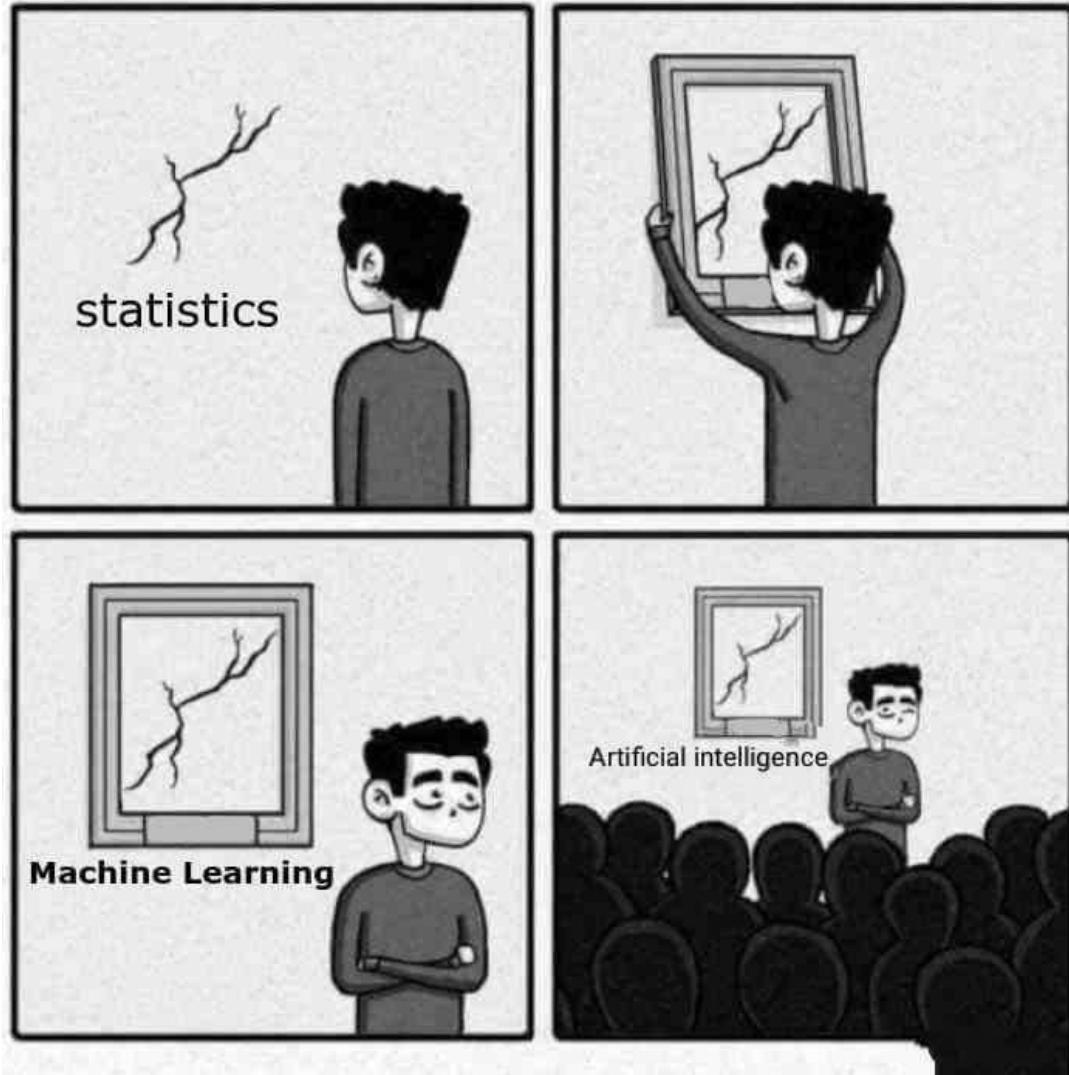
2000's

2010s

ORACLE®

Copyright © 2018, Oracle and/or its affiliates. All rights reserved. |

Estadística, Machine & Deep Learning e Inteligencia Artificial



Índice

1

Introducción

2

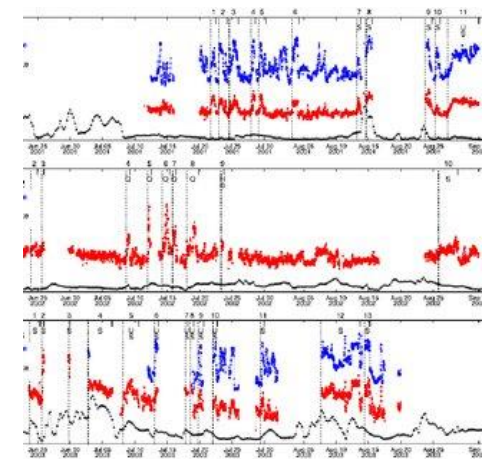
Estadística, Machine Learning, Deep Learning e Inteligencia Artificial

3

¿Por qué utilizar RNA en series de tiempo?

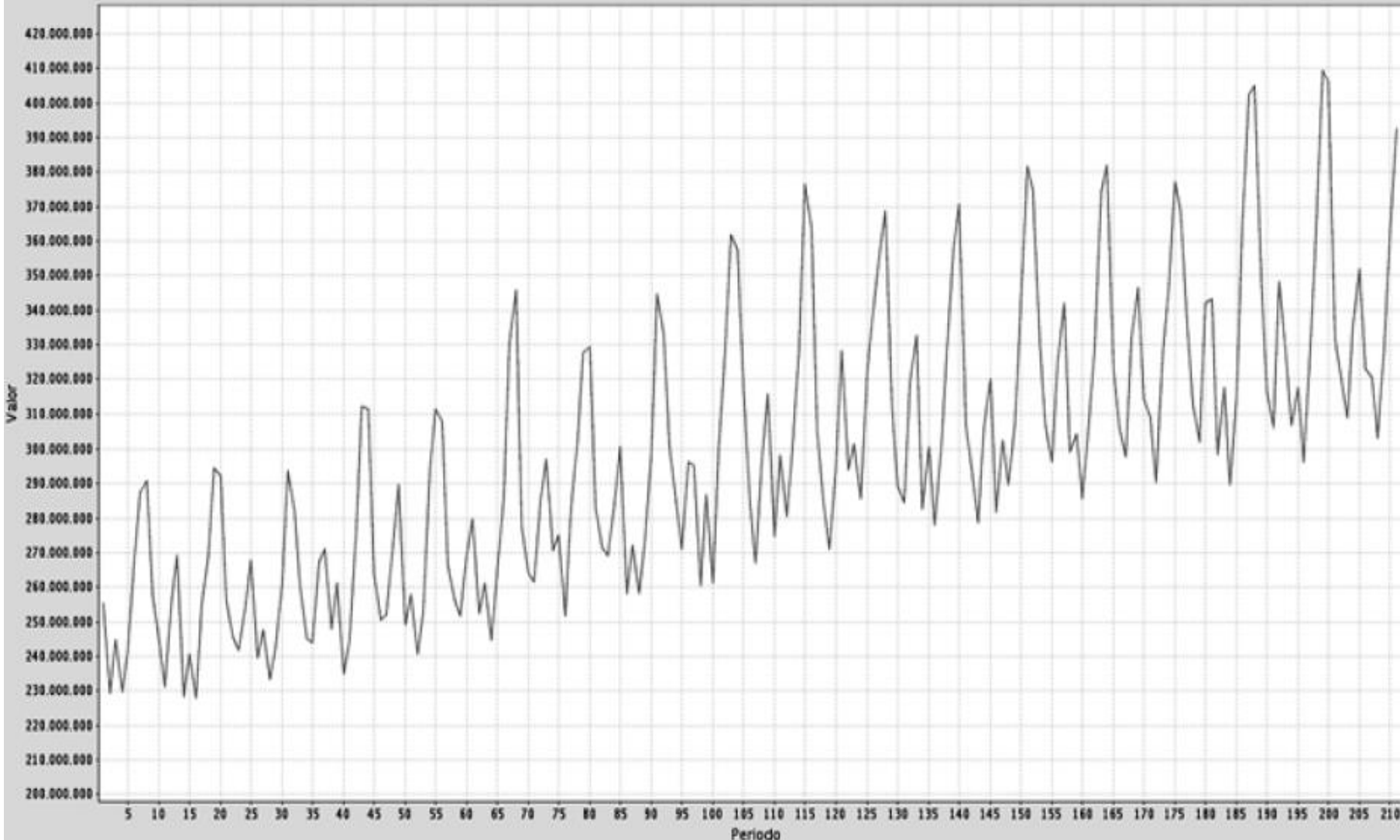
La dificultad de los modelos paramétricos

- Hasta ahora hemos visto únicamente el enfoque paramétrico en la modelización de las series de tiempo: regresión, modelos exponenciales, y el ARIMA.
- Los modelos paramétricos poseen ciertas falencias: difícil identificación, incorrecta estimación de los parámetros, sobre parametrización, falta de ajuste, poca adecuación si no hay linealidad, etc. En sí no se cumple que los residuos son un ruido blanco.
- Al ser las redes neuronales un modelo no paramétrico, utilizar funciones de impulso que aproximan mejor los valores esperados, y al utilizar otra metodología en el proceso de estimación (método no analítico, sino iterativo), los métodos de estimación son, a veces, superiores a los vistos. ¿Qué quiero decir con lo anterior?



La dificultad de los modelos paramétricos

Generación de Electricidad Mensual

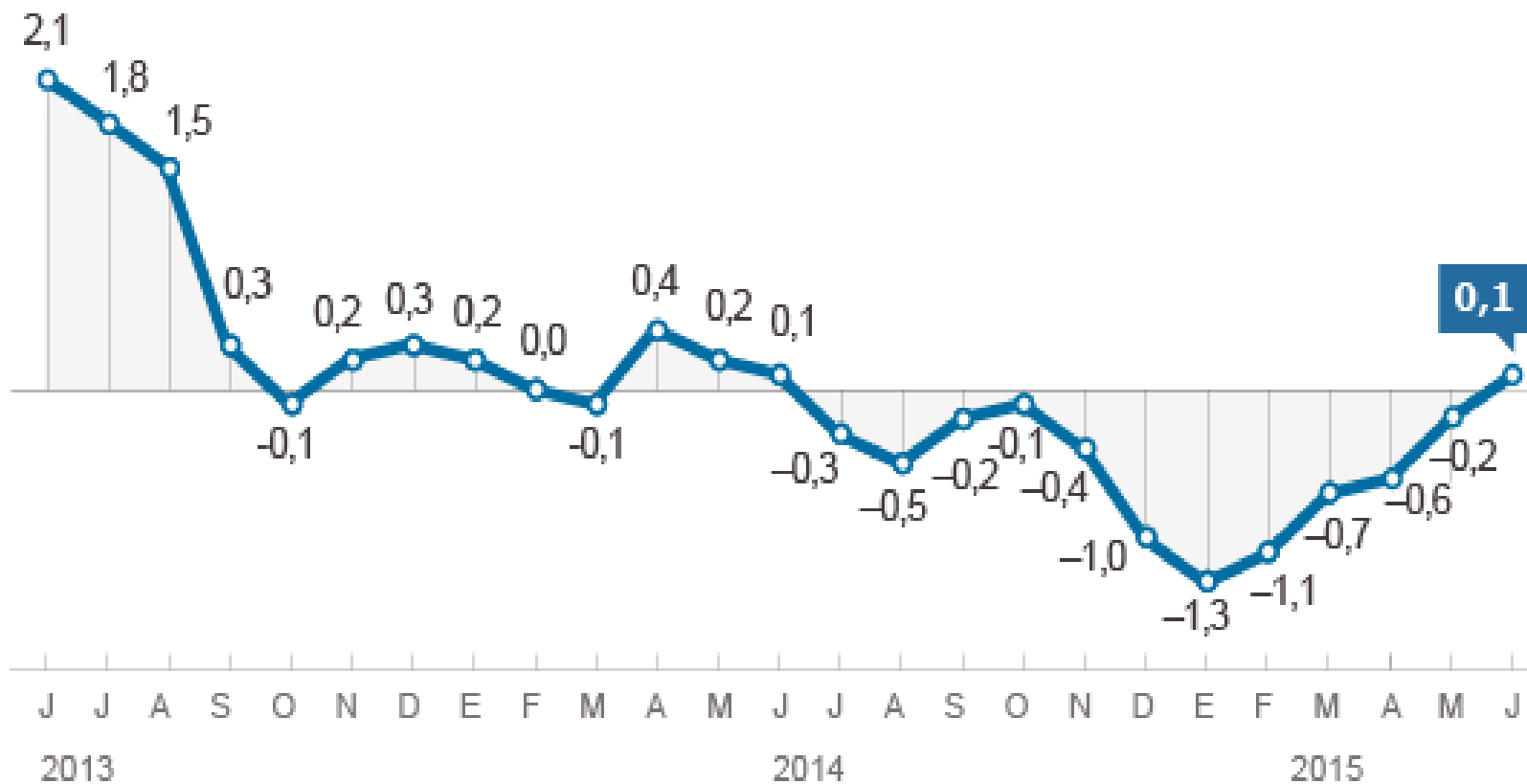


Captación inexacta de la estacionalidad

La dificultad de los modelos paramétricos

EVOLUCIÓN DEL IPC

Variación anual en %

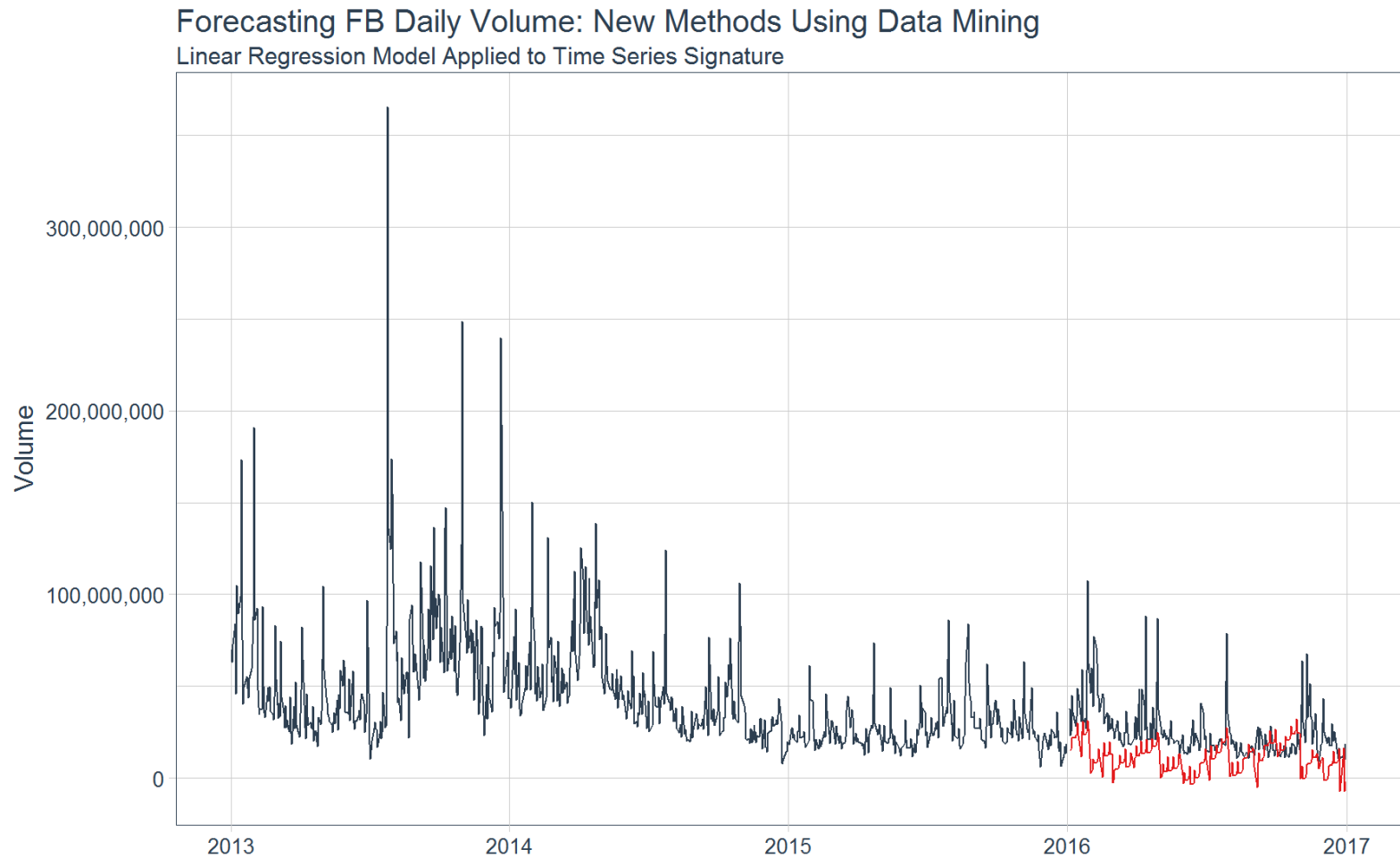


Cambios estructurales debido a variables macro económicas

Fuente: Instituto Nacional de Estadística (INE).

La dificultad de los modelos paramétricos

Series financiera: NN5-001 → una alta volatilidad en los datos...



Data from Yahoo! Finance: 'FB' Daily Volume from 2013 to 2016.

Alto volatilidad en datos

Índice

1

Introducción

2

¿Por qué utilizar RNA en series de tiempo?

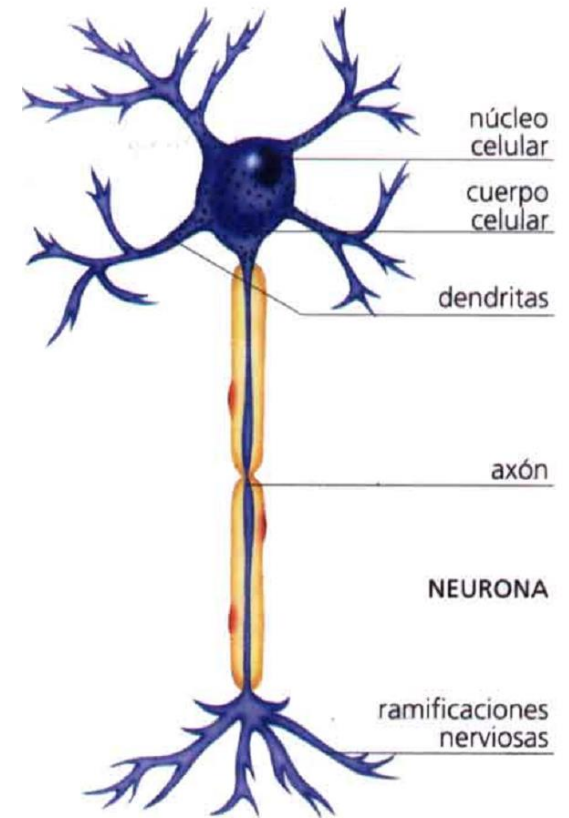
3

Neurona biológica vs neurona artificial

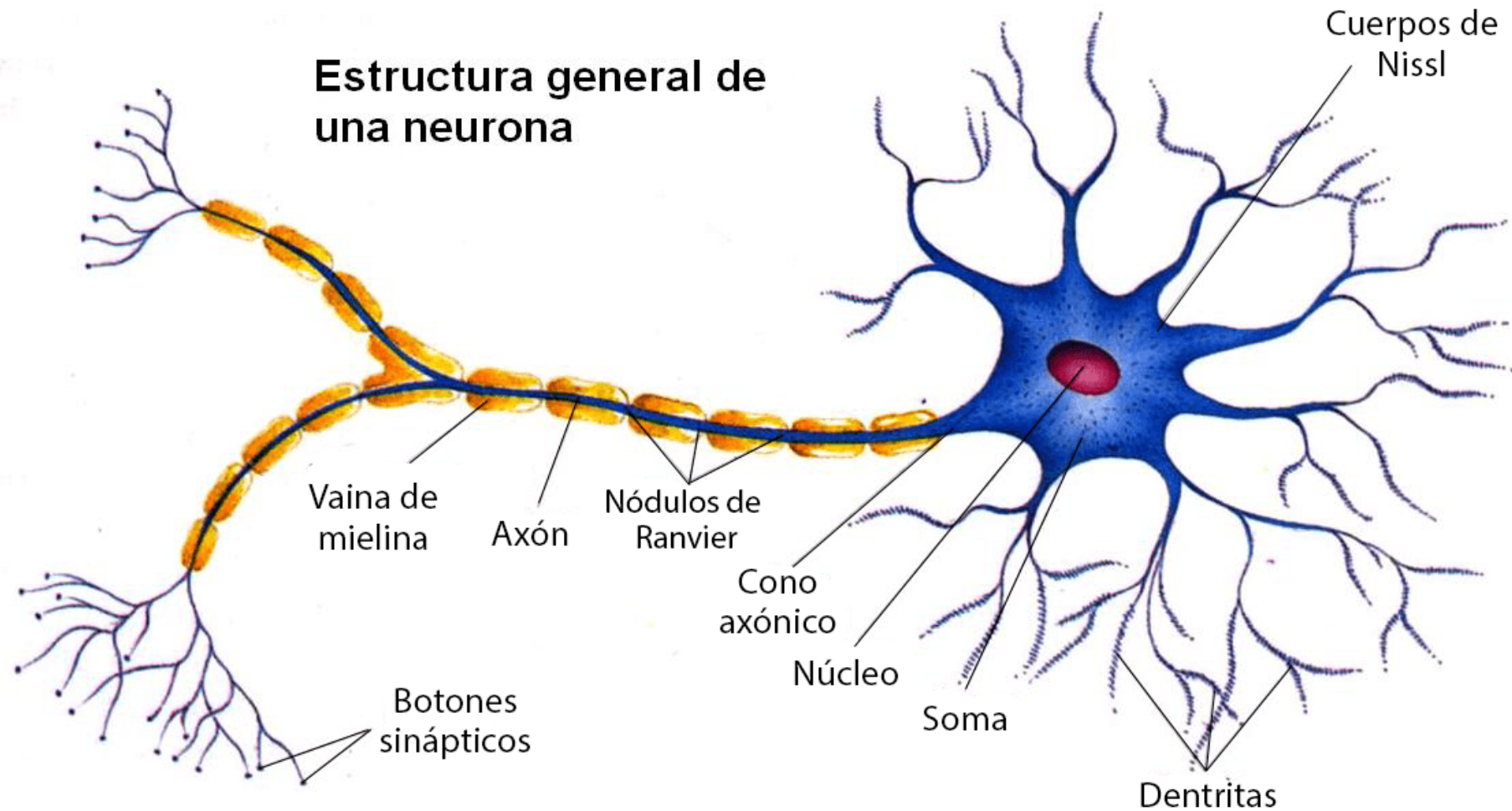
Neurona: biológica vs artificial

Neurona biológica

- Las neuronas biológicas son células nerviosas que constituyen los elementos primordiales del sistema nervioso central.
- Una neurona es capaz de recibir información desde miles de otras neuronas, procesarla y luego generar una nueva información que enviará a otras neuronas con las que está conectada.
- Se estima que el cerebro está compuesto por más de diez billones de neuronas y que cada una está conectada a más de diez mil neuronas. Una neurona biológica está compuesta de: cuerpo celular o soma, axón y múltiples dendritas.

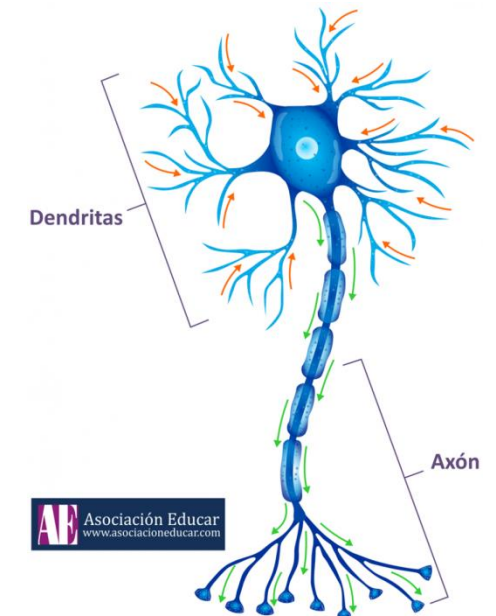
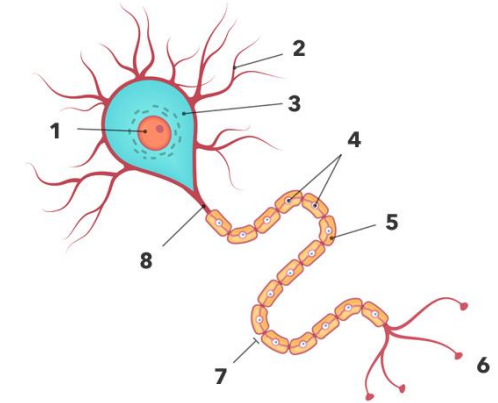


Neurona: biológica vs artificial



Neurona: biológica vs artificial

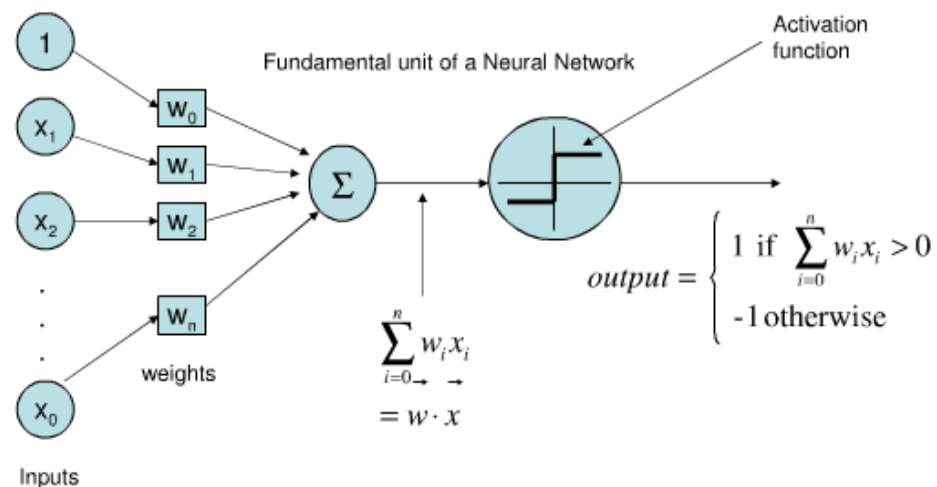
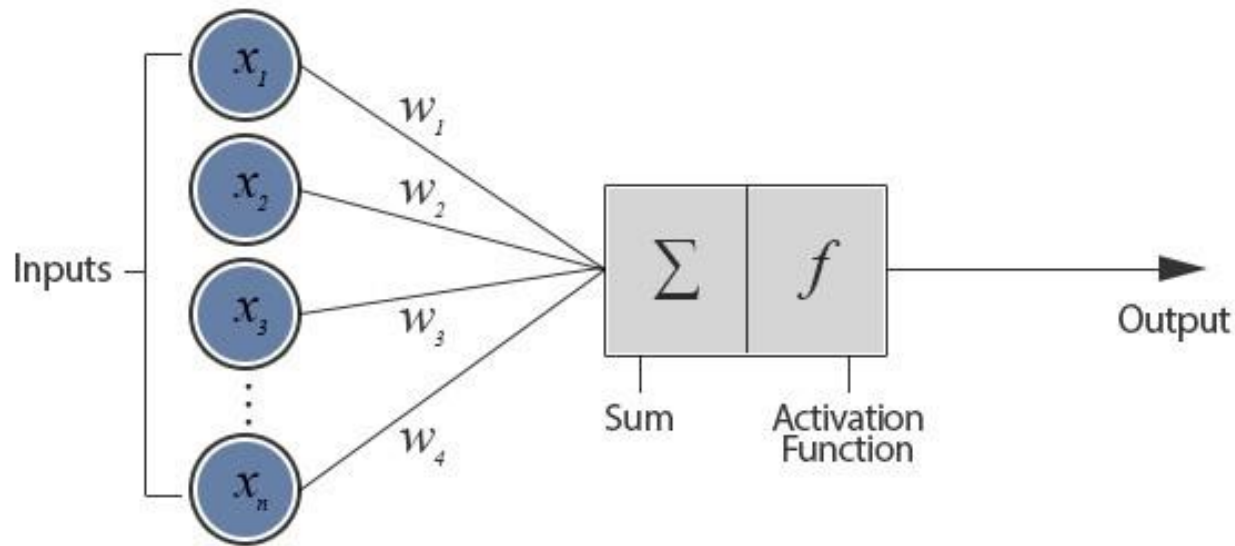
- Las dendritas actúan como elementos receptores que captan las señales provenientes desde otras neuronas y las llevan al cuerpo celular.
- El axón actúa como elemento de salida por el cual se envía la información desde el cuerpo celular a otras neuronas mediante la sinapsis, que es la conexión entre las neuronas en la que se produce un intercambio químico que puede producir la excitación o inhibición de la neurona receptora.
- El cuerpo celular de la neurona es donde se realiza el procesamiento de las señales entrantes.



Neurona: biológica vs artificial

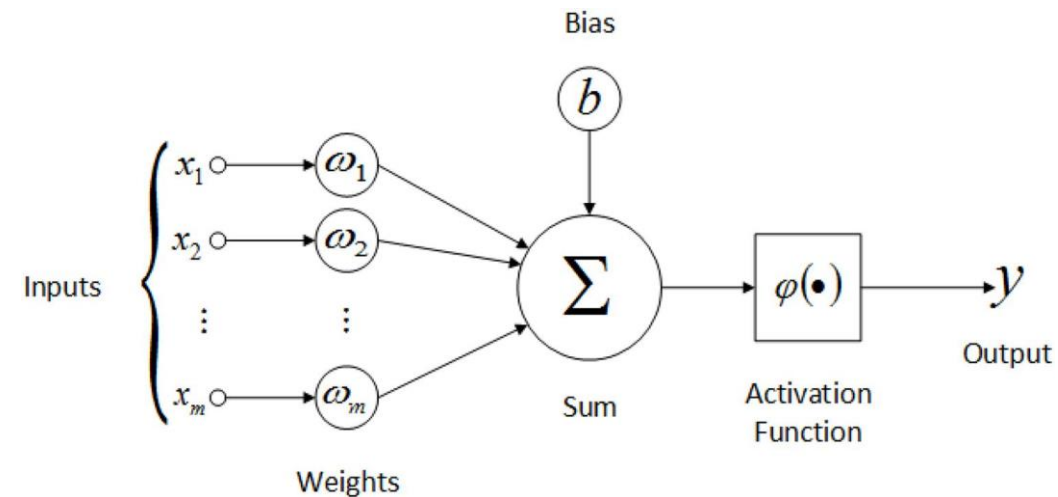
Neurona Artificial

Una neurona artificial o procesador elemental es una abstracción de una neurona biológica que opera como una unidad de procesamiento de información que es fundamental para la operación de una red neuronal. Una neurona artificial está compuesta de: un conjunto de entradas, un conjunto de pesos sinápticos, un “cuerpo celular” y una salida.

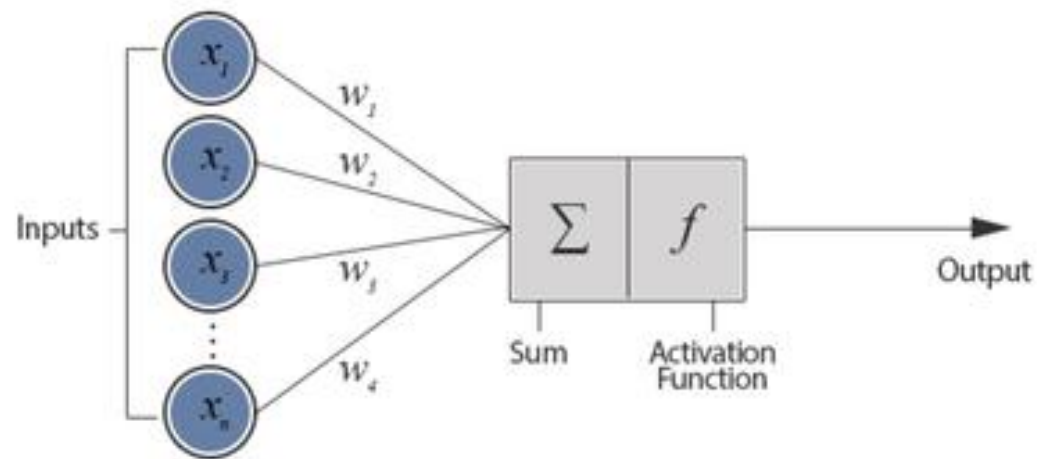
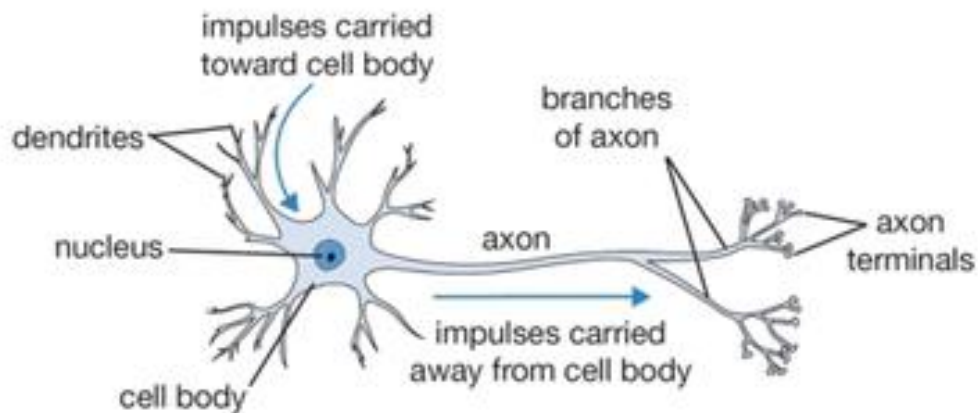


Neurona: biológica vs artificial

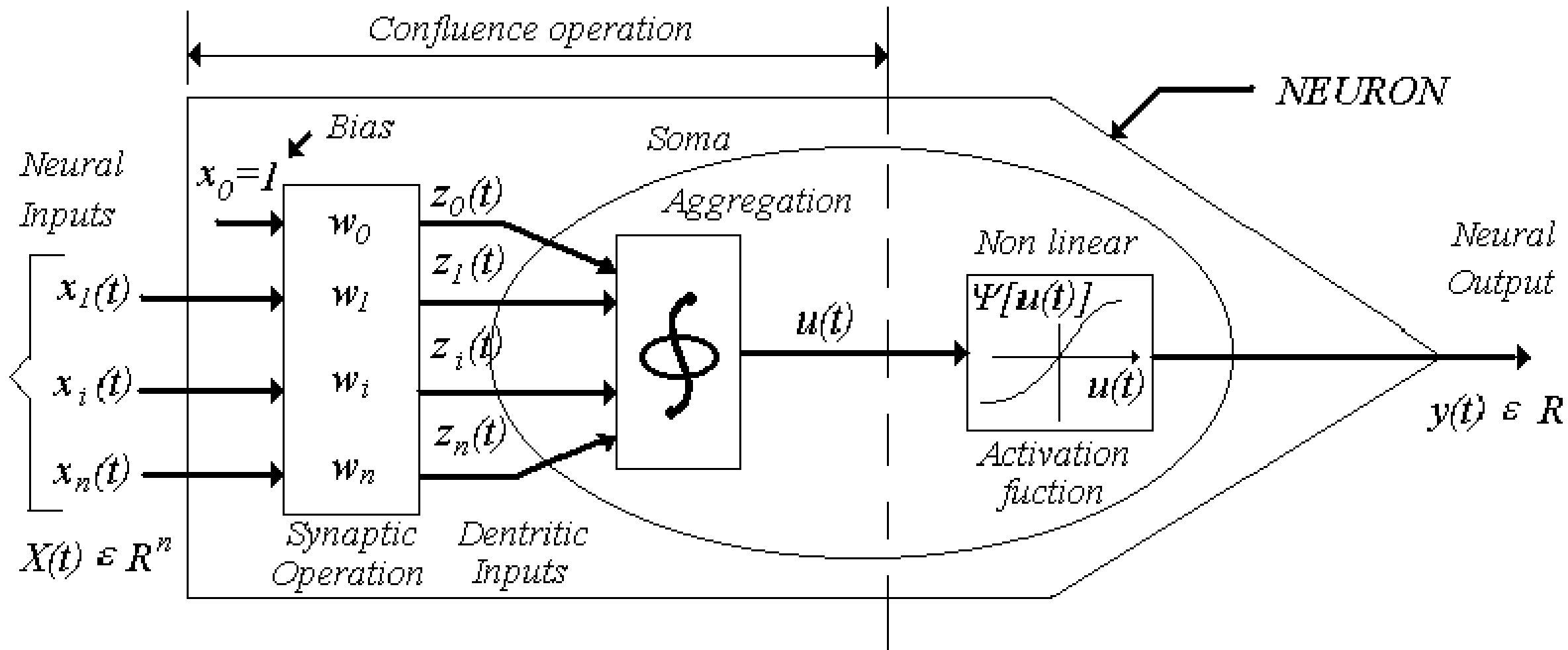
La neurona artificial tratará de simular la funcionalidad de la neuronal biológica.



Biological Neuron versus Artificial Neural Network

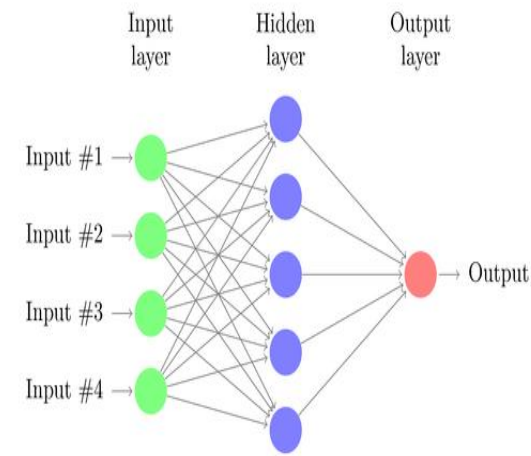


Neurona: biológica vs artificial

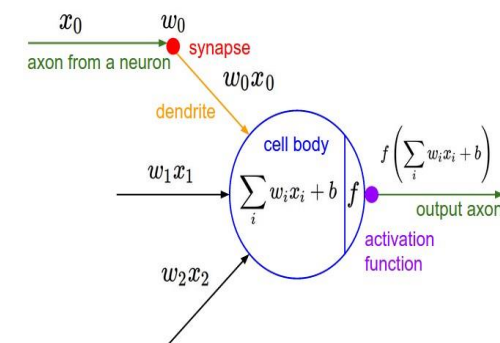


Neurona: biológica vs artificial

El conjunto de entradas corresponde a los valores que son presentados a la neurona como si se tratase de las dendritas en la versión biológica (las variables). Los pesos sinápticos corresponden a las ponderaciones de cada entrada y representan en un modelo neuronal artificial el conocimiento de la neurona. La salida corresponde a la información que es enviada desde la neurona hacia las otras neuronas con las cuales existe sinapsis, como si se tratase del axón en la versión biológica.



El “cuerpo celular” es el encargado de realizar el procesamiento de la información que proviene del conjunto de entradas y entregarla a la salida. El “cuerpo celular” está compuesto de una función de red y una función de activación.

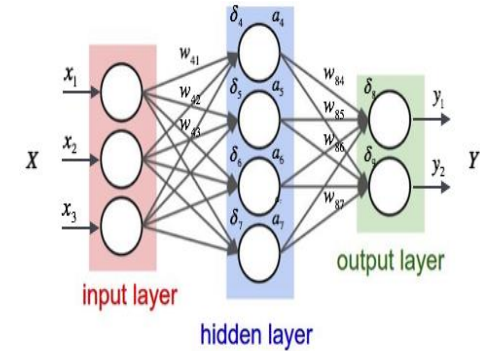


Y esto en lenguaje de estimación...

Neurona: biológica vs artificial

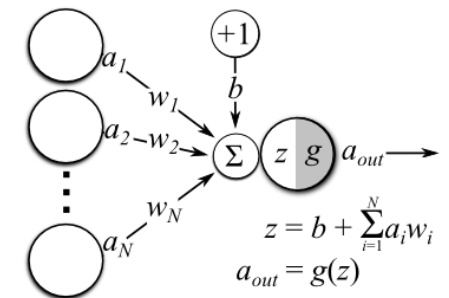
Pesos sinápticos

Los pesos sinápticos son un conjunto de valores asociados con cada una de las entradas, y son ajustados mediante el uso de un algoritmo de aprendizaje con el objetivo de plasmar en ellos el conocimiento de un problema determinado. Los pesos sinápticos se representan matemáticamente con un vector W , donde w_i es el peso correspondiente a la entrada i -ésima de la neurona.



Función de red

La función de red, también conocida como función de propagación, calcula el valor base en la neurona de acuerdo al conjunto de entradas y pesos sinápticos relacionados, siendo la más utilizada la Función de Base Lineal (FBL), que consiste en la sumatoria de las entradas ponderadas con los pesos sinápticos.



Neurona: biológica vs artificial

- La función de red FBL para un procesador elemental j al que están conectadas n entradas se puede definir formalmente como:

$$F(X, W_j) = \sum_{i=1}^n x_i * w_{ij}$$

donde X es el vector de entradas, W_j es el vector de pesos sinápticos que une a las entradas con el procesador elemental e i es el índice que identifica a cada entrada desde 1 hasta n .

- Otra función de red es la utilizada en Redes Función de Base Radial (FBR), y consiste en determinar el valor de red calculando la distancia a un determinado punto de referencia. La función de red FBR para un procesador elemental j al que están conectadas n entradas se puede definir formalmente como:

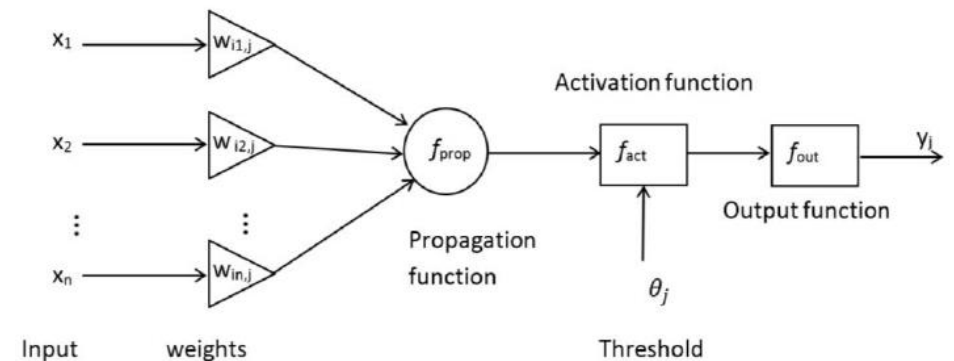
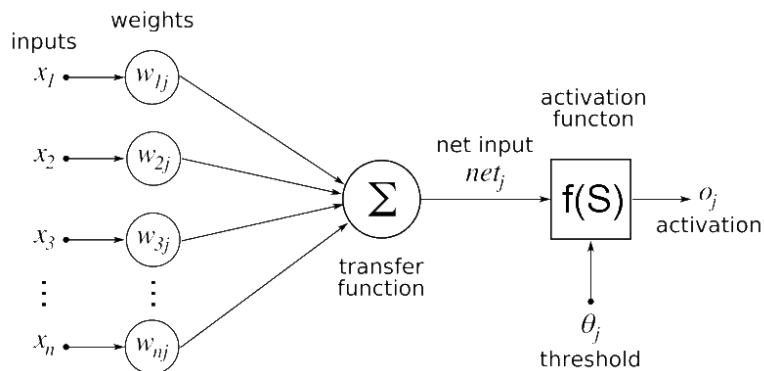
$$F(X, W_j) = \sqrt{\sum_{i=1}^n (x_i - w_{ij})^2}$$

Neurona: biológica vs artificial

Función de activación

La función de activación, también conocida como función de salida, es considerada como una de las características más importantes de una neurona artificial, dado que representa el comportamiento de acuerdo al resultado.

- Esta función es la encargada de calcular el estado de activación de la neurona en base al resultado entregado por la función de red.
- Generalmente los resultados entregados por la función de activación se encuentran en el intervalo $[-1, 1]$ o $[0, 1]$, donde un estado de actividad completa es representado por el valor 1 y un estado de inactividad completa por 0 o -1. Se suelen utilizar funciones trigonométricas como seno, cosenos, tagente, etc...



Índice

1

Introducción

4

Redes Neuronales
Artificiales

2

¿Por qué utilizar RNA en
series de tiempo?

3

Neurona biológica vs
neurona artificial

Redes Neuronales Artificiales

- Una Red Neuronal Artificial es un paradigma de procesamiento de la información que es inspirado en el modo en que un sistema nervioso biológico, como el cerebro, procesa la información. Una RNA está compuesta por un conjunto de neuronas artificiales o procesadores elementales, los cuales, interconectados de alguna manera, trabajan para obtener la solución de un problema específico. En la literatura existen diversas definiciones para una RNA, pero una de las más certeras con el propósito de resumir su todo es la siguiente: “Una Red Neuronal es un conjunto de procesadores elementales interconectados, no lineal *ni estacionario*, que realiza al menos alguna de las siguientes funciones: Aprendizaje, Memorización, Generalización o Abstracción de características esenciales, a partir de un conjunto de ejemplos”.

- Las neuronas artificiales o nodos pueden clasificarse, de acuerdo a su funcionalidad, de la siguiente forma:

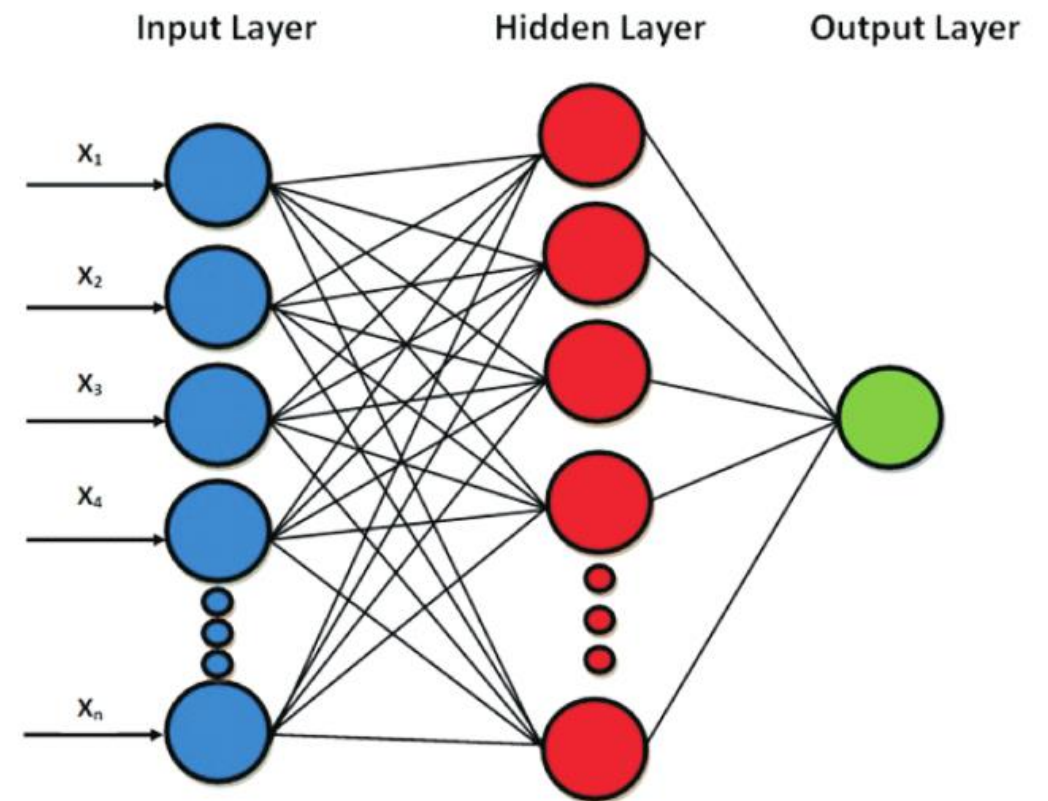
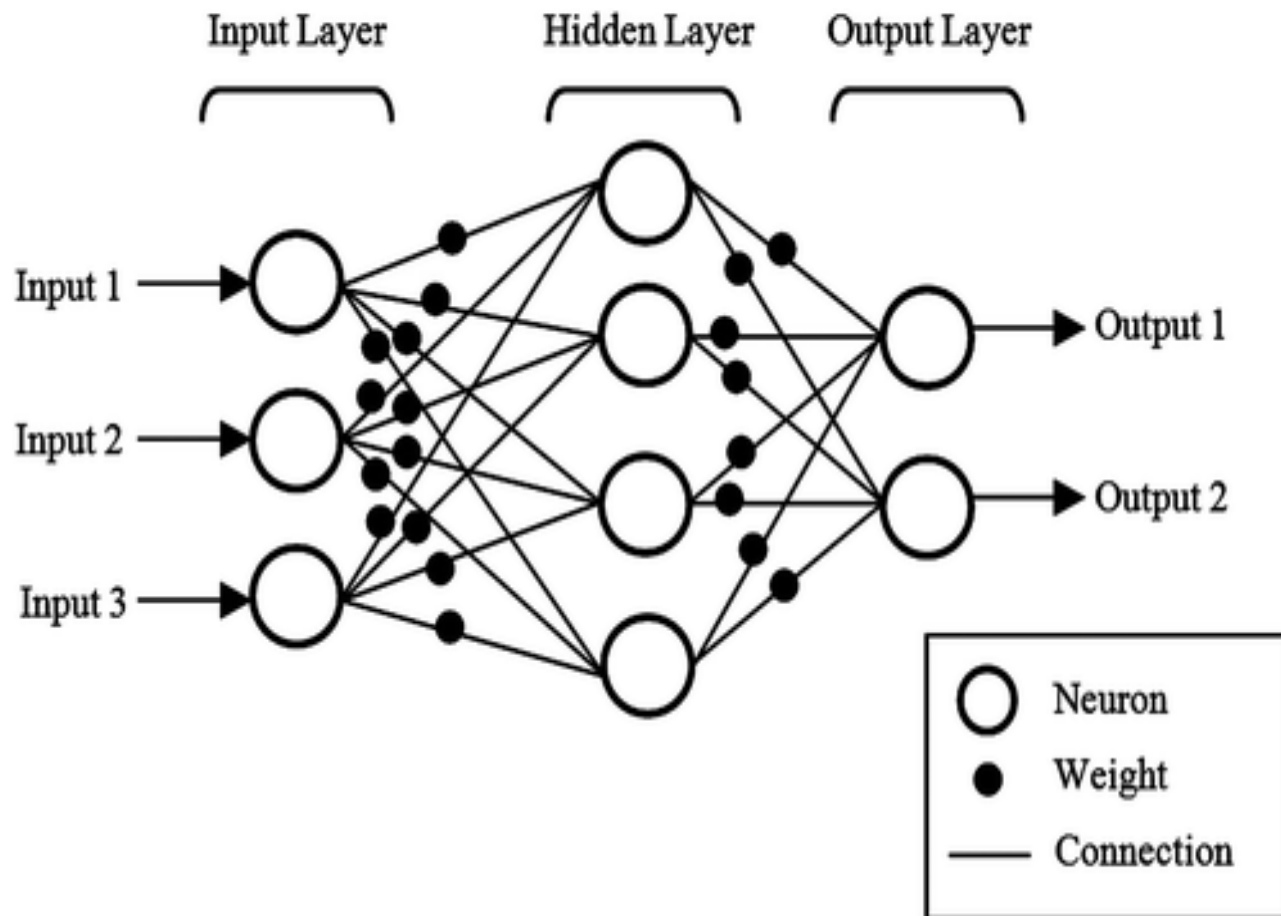
Nodos de entrada: son los encargados de recibir las señales o informaciones desde el exterior de la red. No realizan procesamiento de la información.

Nodos salida: son los encargados de enviar la salida de la red neuronal hacia el exterior de ésta. Pueden realizar procesamiento de la información.

Nodos ocultos: son los encargados de realizar el procesamiento de la información de la información y, con esto, realizar el aprendizaje. Reciben la información desde los nodos de entrada y la envían a los nodos de salida.

Redes Neuronales Artificiales

Se conoce como capa o nivel a un conjunto de neuronas del mismo tipo, ya sea de entrada, salida u oculta. Toda RNA incorpora una capa de entrada, una capa de salida y una o más capas ocultas.



Redes Neuronales Artificiales

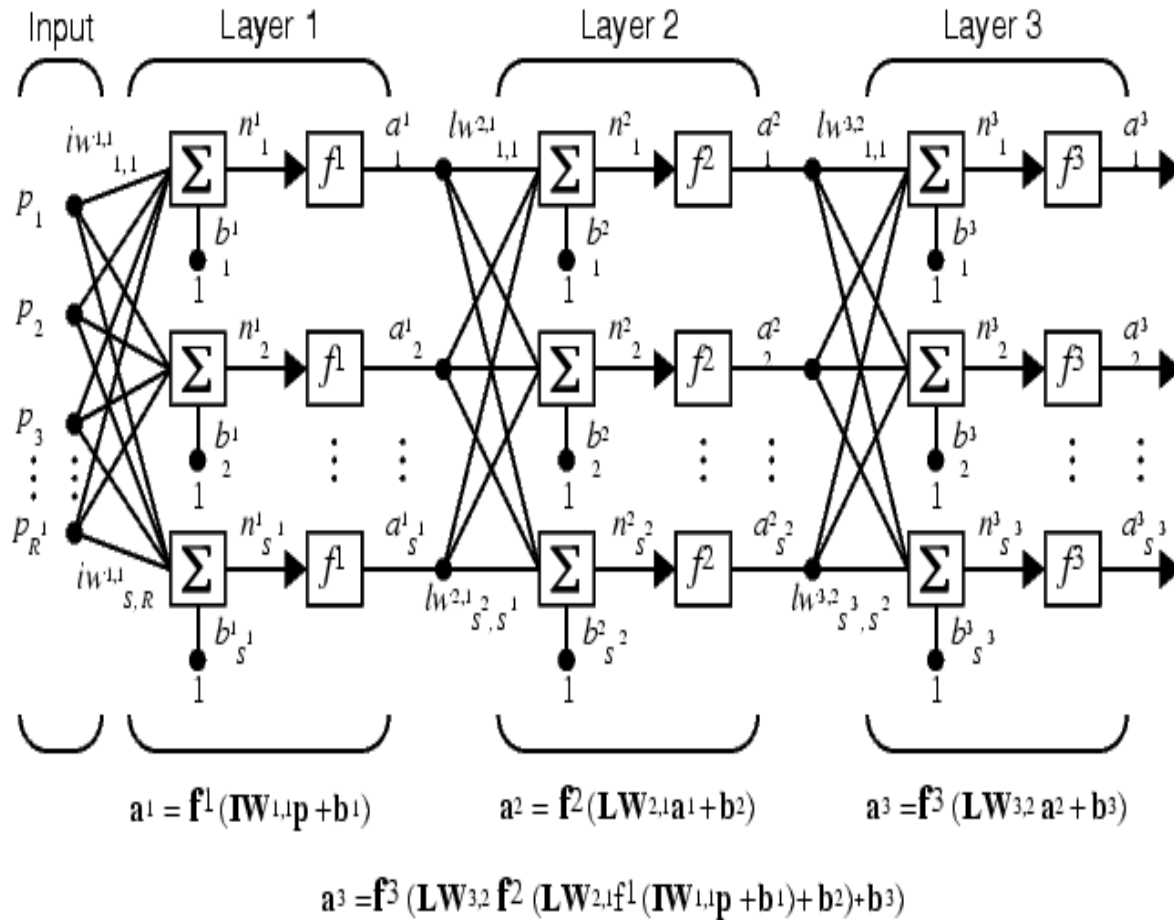
En el estudio de las RNA deben considerarse tres aspectos fundamentales: **la arquitectura, el aprendizaje y la capacidad de generalización de la red.**

a) Arquitectura de red

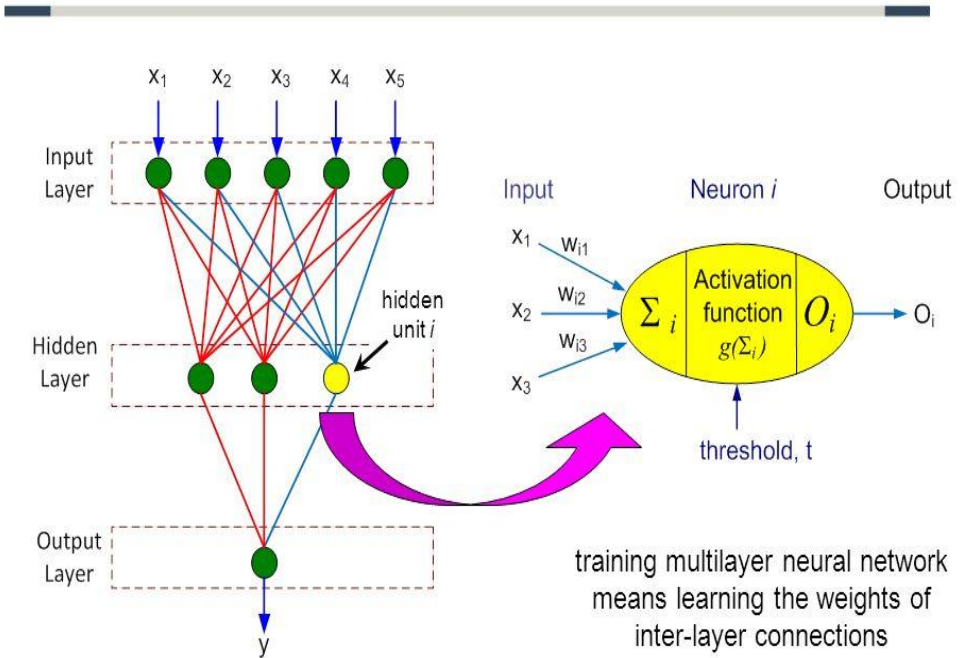
El término arquitectura de la red está relacionado con el diseño estructural de la red y busca determinar los siguientes elementos: la cantidad de entradas y salidas, la cantidad de nodos ocultos, la función de red y de activación asociada a cada nodo, la forma en que los nodos están interconectados, la dirección que sigue la información, y la selección de un conjunto de datos adecuado para realizar el entrenamiento y la validación del modelo obtenido.

Redes Neuronales Artificiales

Un ejemplo de una arquitectura, podría ser la siguiente:



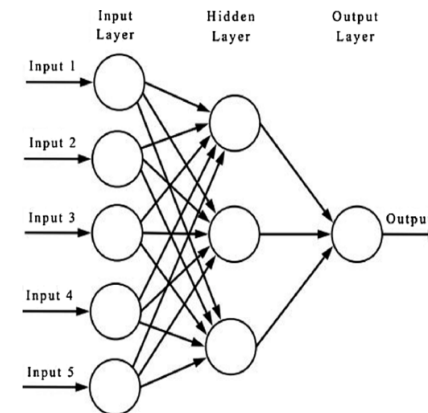
General structure of multilayer neural network



Redes Neuronales Artificiales

Cantidad de nodos de entrada y salida

- La determinación del conjunto de datos de entrada y de salida de la red depende exclusivamente del problema que se busca resolver. La cantidad de entradas es una labor basta más sencilla... es identificar las variables que queremos introducir al modelo. El problema serían los nodos intermediarios y de salida.
- Es una práctica común realizar un análisis de los datos tanto de entradas como de salida, para escalarlo y/o pre procesarlos. Por lo general, el escalamiento se realiza en el intervalo $[0, 1]$ o $[-1, 1]$ y el pre-procesamiento consiste en aplicar algún método estadístico que permita capturar la información relevante de los datos de entrada y desechar la restante.
- Si se ha realizado un escalamiento de los datos de entrada, se debe realizar una operación de desnormalización con los datos de salida para volver a su estado original los datos que fueron escalados.



Feature Scaling Formula

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

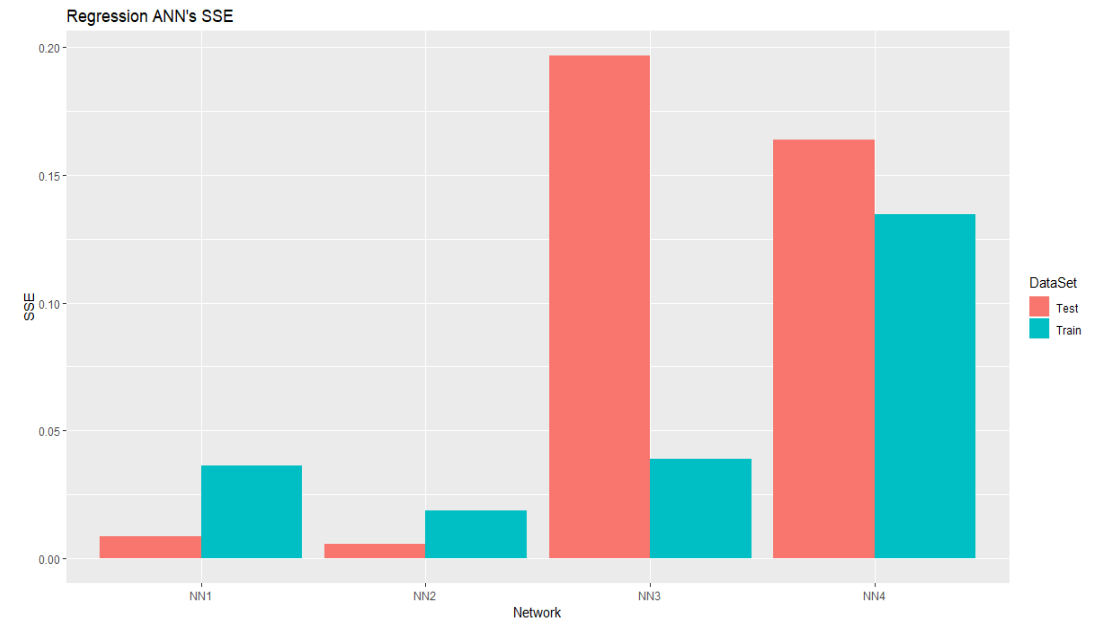
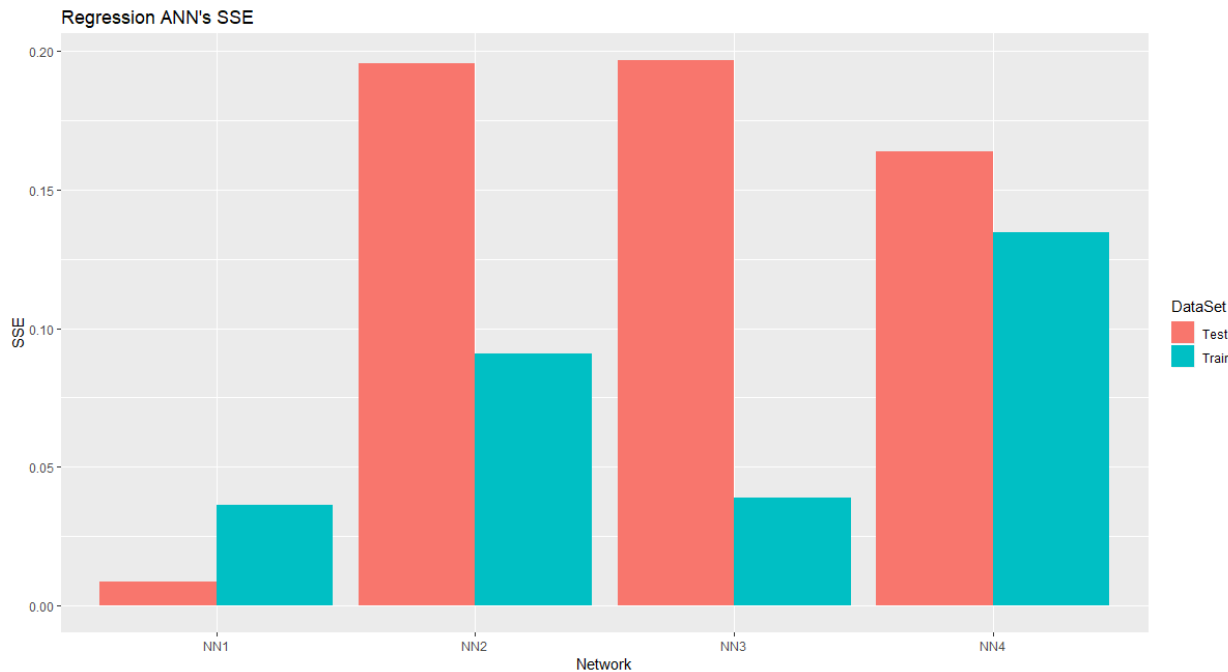
new (rescaled) feature \leftarrow X'
old feature(s) \leftarrow X
115 \leftarrow X_{\min}
175 \leftarrow X_{\max}
info taken from old feature(s) \leftarrow X

Quiz
old weights: [115, 140, 175]
What is X'_{140} ? 5/12

Redes Neuronales Artificiales

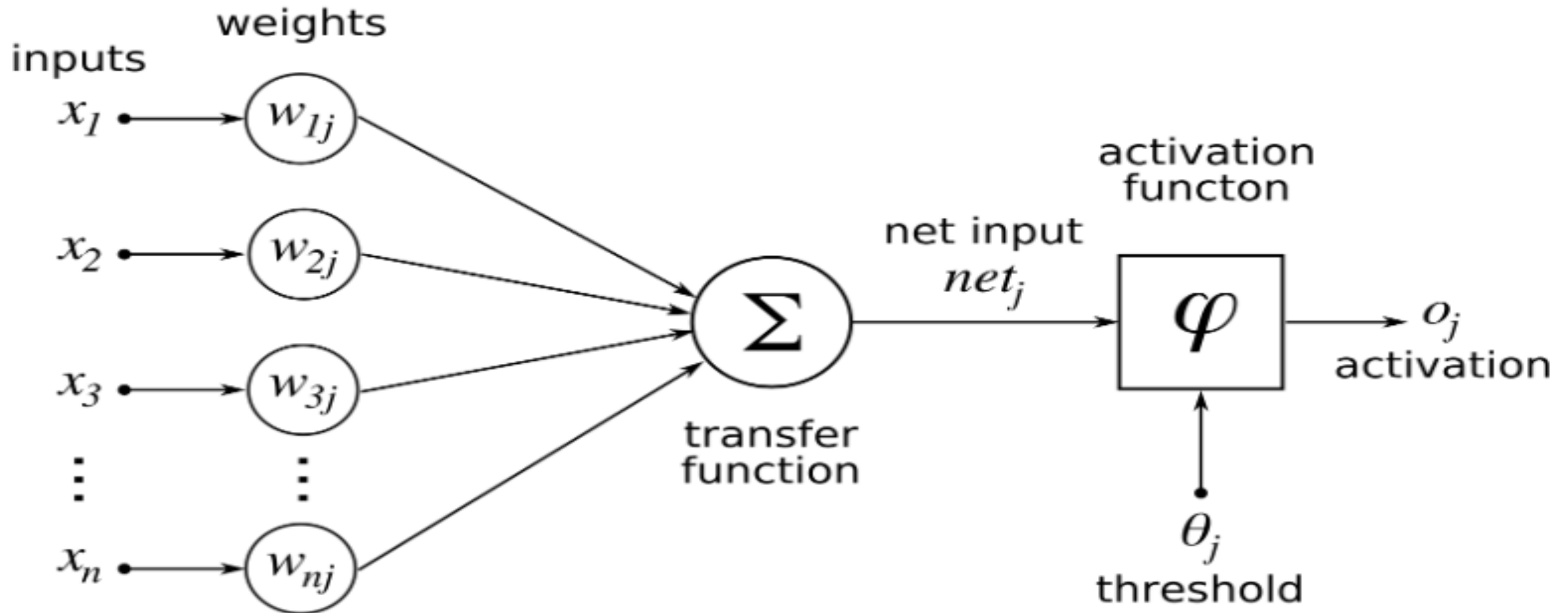
Cantidad de capas ocultas

La cantidad de capas ocultas se determina de acuerdo a la complejidad del problema. En la mayoría de los problemas basta con utilizar una sola capa oculta, pero si no se consiguen buenos resultados puede intentarse utilizando más de una. La cantidad de nodos para una capa cualquiera se determina por lo general por ensayo y error, variando la cantidad de nodos entre el $n/2$ y $2n$, por lo general, siendo n el número de nodos de la capa anterior.



Redes Neuronales Artificiales

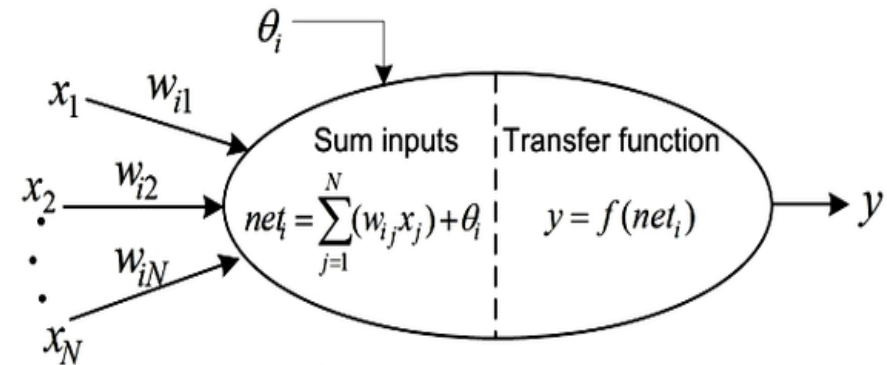
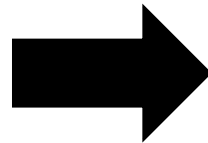
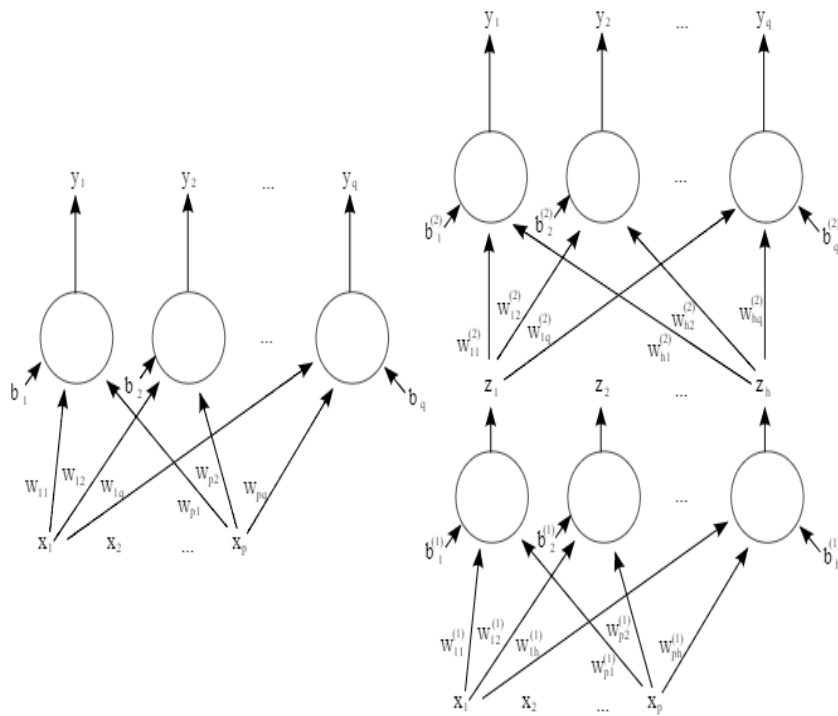
Veamos ahora dos funciones imprescindibles de las RNA: función de red o transferencia y la función de activación.



Redes Neuronales Artificiales

Función de red o función de transferencia

La función de red o de propagación es estándar para la gran mayoría de los problemas. En la gran mayoría de casos, se utiliza la Función de Base Lineal, que consiste en las sumas de las entradas ponderadas con los pesos sinápticos correspondientes.



$x_i (i = 1, 2, \dots, N)$: Set of inputs

w_{ij} : Connection weights

θ_i : Bias term

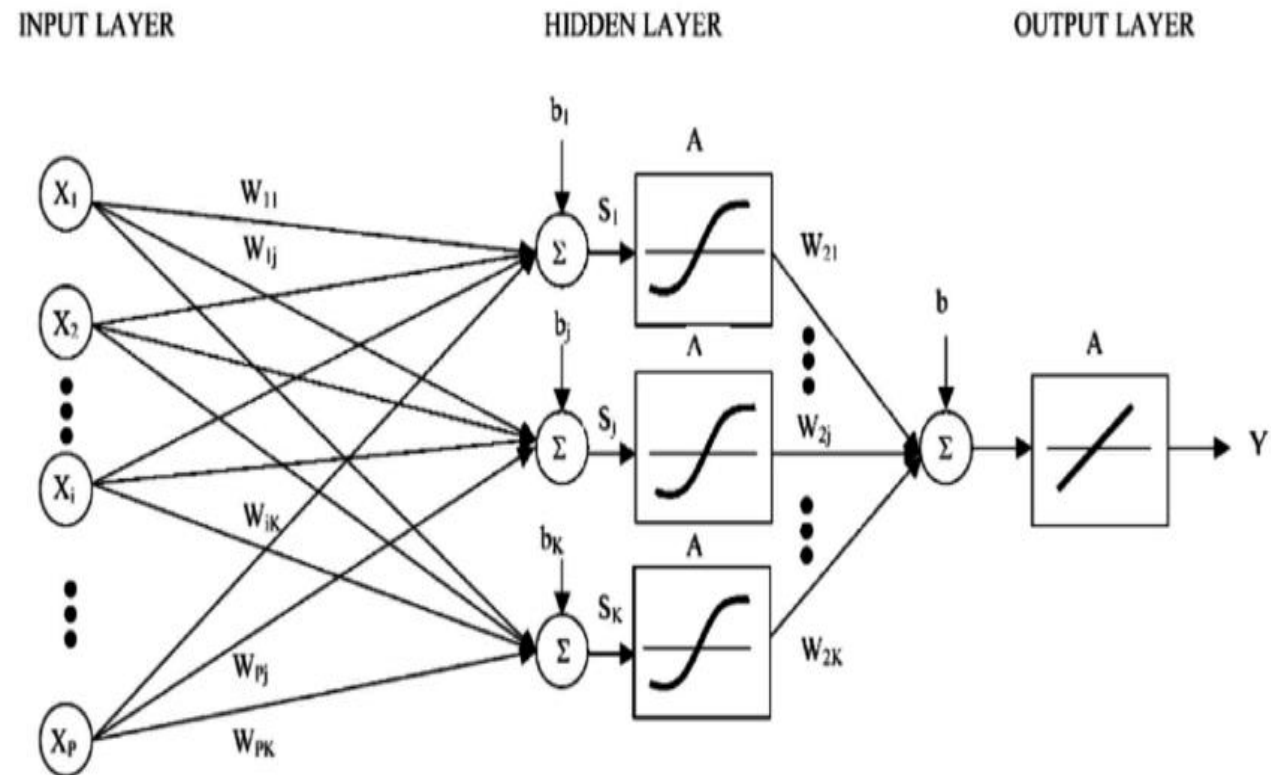
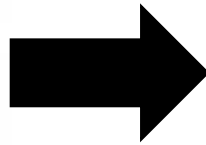
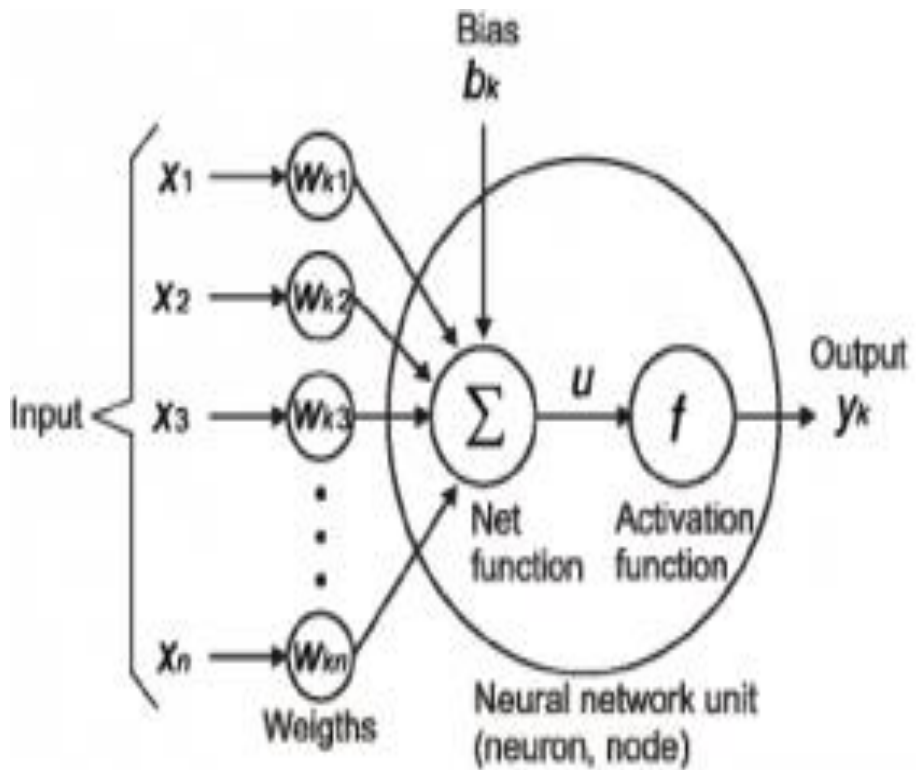
Net_i : Net input signal

Transfer function : tangent sigmoid function $f(x) = \left[\frac{2}{1 + e^{-2x}} - 1 \right]$ or linear function $f(x) = x$

Redes Neuronales Artificiales

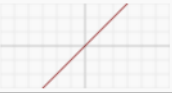




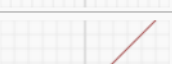



Función de activación

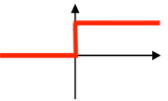
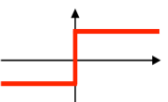
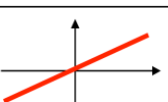
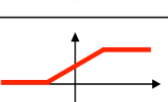
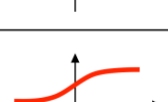

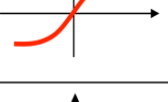

La función de activación es aquella que inversamente a recibir, debe propagar o esparcir la información. Hay una enorme gama de funciones de activación en las capas de entrada: la función Sigmoide o la Tangente Hiperbólica etc, etc, etc..



Redes Neuronales Artificiales

Algunas funciones de activación...

Name	Plot	Equation	Derivative
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a Soft step)		$f(x) = \frac{1}{1 + e^{-x}}$	$f'(x) = f(x)(1 - f(x))$
Tanh		$f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2 + 1}$
Rectified Linear Unit (ReLU)		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Parametric Rectified Linear Unit (PReLU) [2]		$f(x) = \begin{cases} \alpha x & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
Exponential Linear Unit (ELU) [3]		$f(x) = \begin{cases} \alpha(e^x - 1) & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} f(x) + \alpha & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$
SoftPlus		$f(x) = \log_e(1 + e^x)$	$f'(x) = \frac{1}{1 + e^{-x}}$

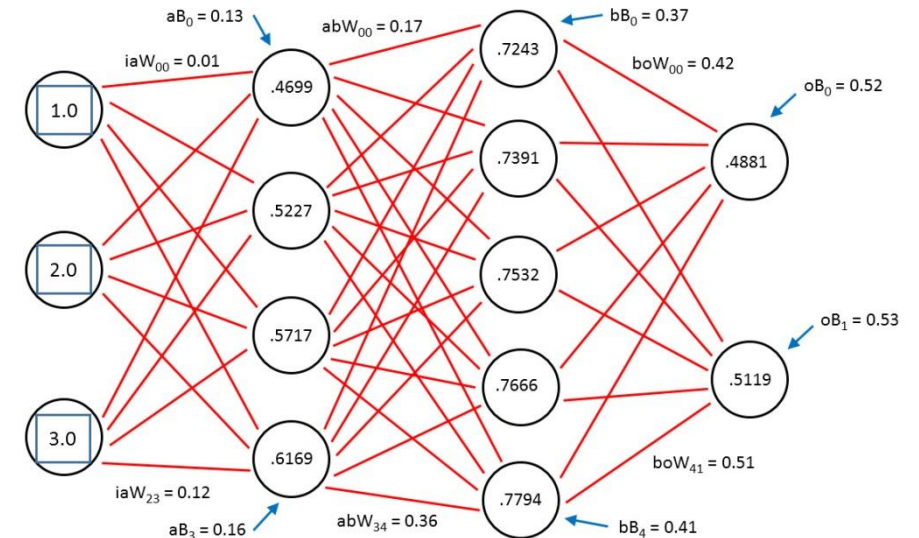
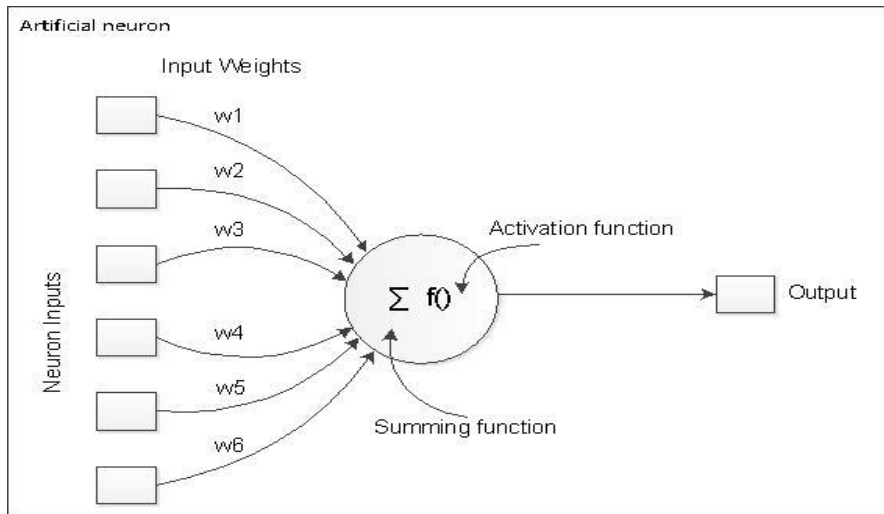
Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = \max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	

Redes Neuronales Artificiales

Interconexión

La interconexión de una red se refiere a la forma en que se producen las conexiones entre las neuronas agrupadas en capas. Se pueden distinguir dos tipos de interconexiones entre capas:

- Totalmente conectadas: las salidas de los nodos de una capa cualquiera están conectadas a todos los nodos de otra capa.
- Parcialmente conectadas: las salidas de los nodos de una capa cualquiera están conectadas a un grupo de nodos de otra capa y no a todos.



Redes Neuronales Artificiales

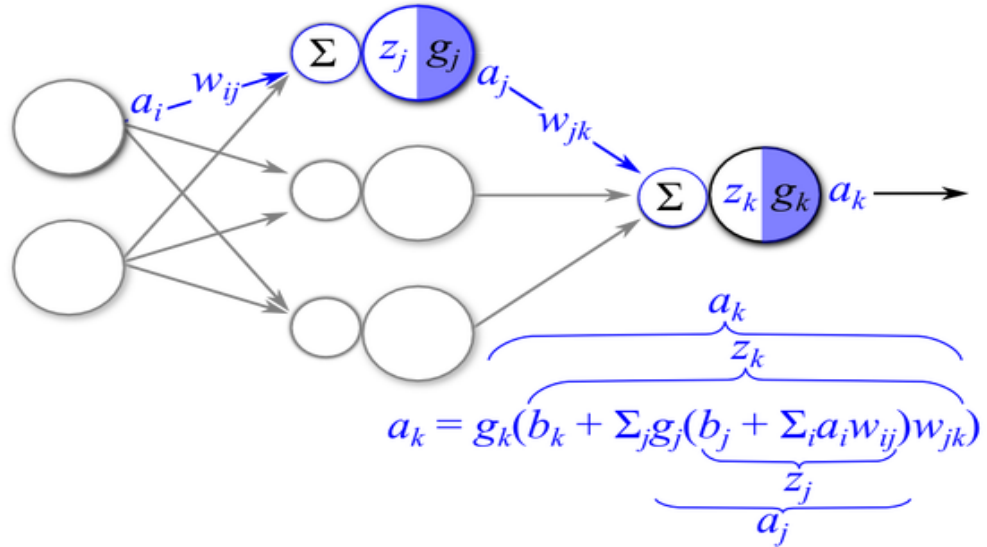
Dirección de la información

La dirección de la información se refiere al flujo que sigue la información en la red neuronal. Se pueden distinguir 4 tipos de redes con respecto a la dirección:

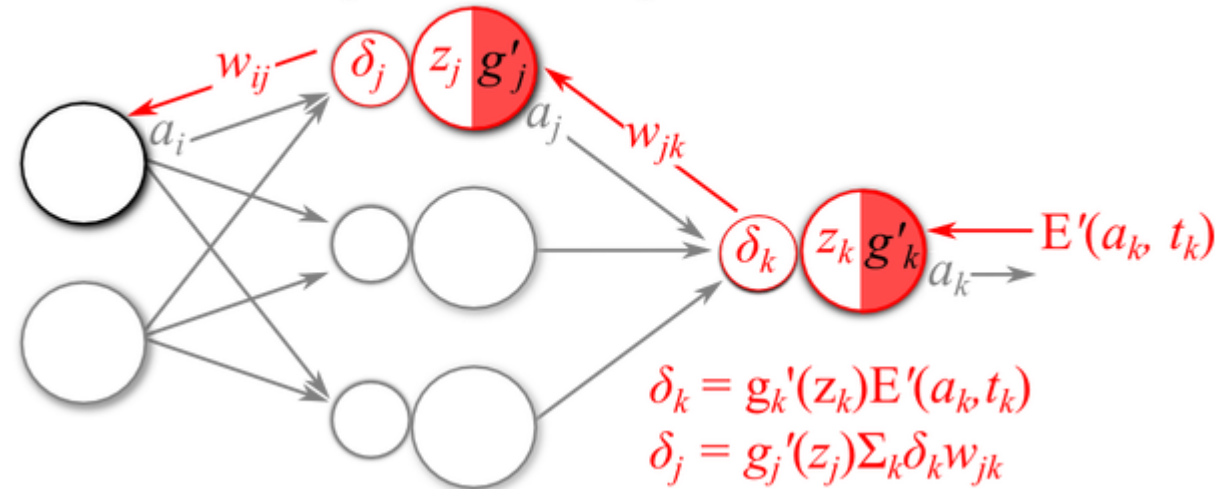
- Redes de alimentación hacia adelante: el flujo de la información siempre es desde una capa a la siguiente, es decir, va desde la capa de entrada, pasa por las capas ocultas para finalizar en la capa de salida. Ejemplo de este tipo de redes es el Perceptrón Multicapa.
- Redes de alimentación hacia atrás: el flujo de la información puede ir desde una capa a capas anteriores. Ejemplo de este tipo de red son las Redes Recurrentes.
- Redes de alimentación lateral: puede existir conexiones entre neuronas de la misma capa, o de distintas capas.
- También, una neurona puede alimentarse por ella misma...

Redes Neuronales Artificiales

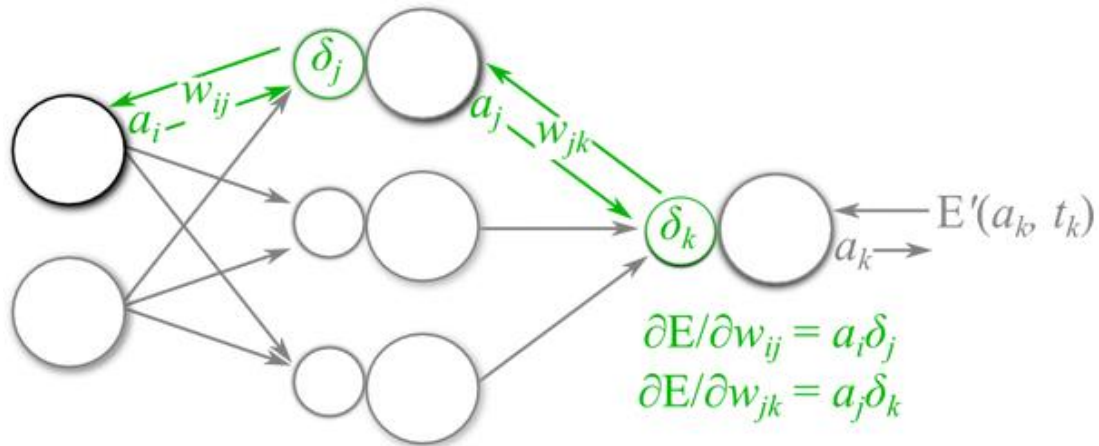
I. Forward-propagate Input Signal



II. Back-propagate Error Signals



III. Calculate Parameter Gradients



IV. Update Parameters

$$w_{ij} = w_{ij} - \eta \left(\frac{\partial E}{\partial w_{ij}} \right)$$

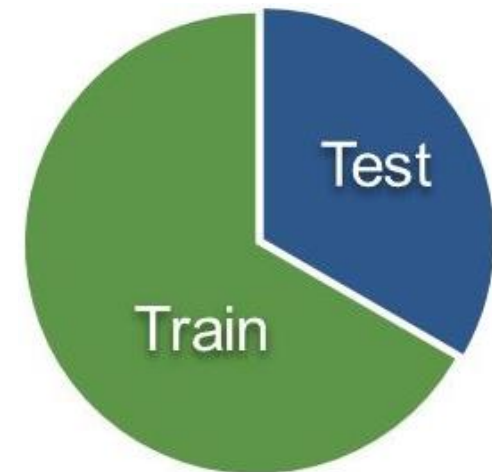
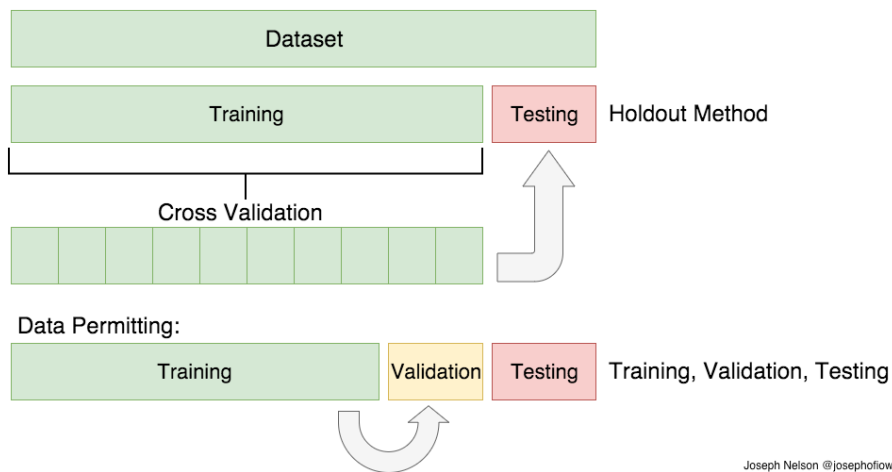
$$w_{jk} = w_{jk} - \eta \left(\frac{\partial E}{\partial w_{jk}} \right)$$

for learning rate η

Redes Neuronales Artificiales

Selección de datos para entrenamiento y validación

Se debe determinar un conjunto de datos que sea completamente representativo del problema a solucionar. Por lo general, se utiliza el 80% de los datos para realizar el entrenamiento de la red, y el 20% restante para realizar la validación del modelo obtenido. De esta forma, se presenta el 80% para que se lleve a cabo el aprendizaje del problema en cuestión y luego se presenta el 20% restante de los datos, para verificar si realmente el modelo entrega resultados aceptables al presentarle patrones que pueden ser desconocidos.



Redes Neuronales Artificiales

b) Aprendizaje

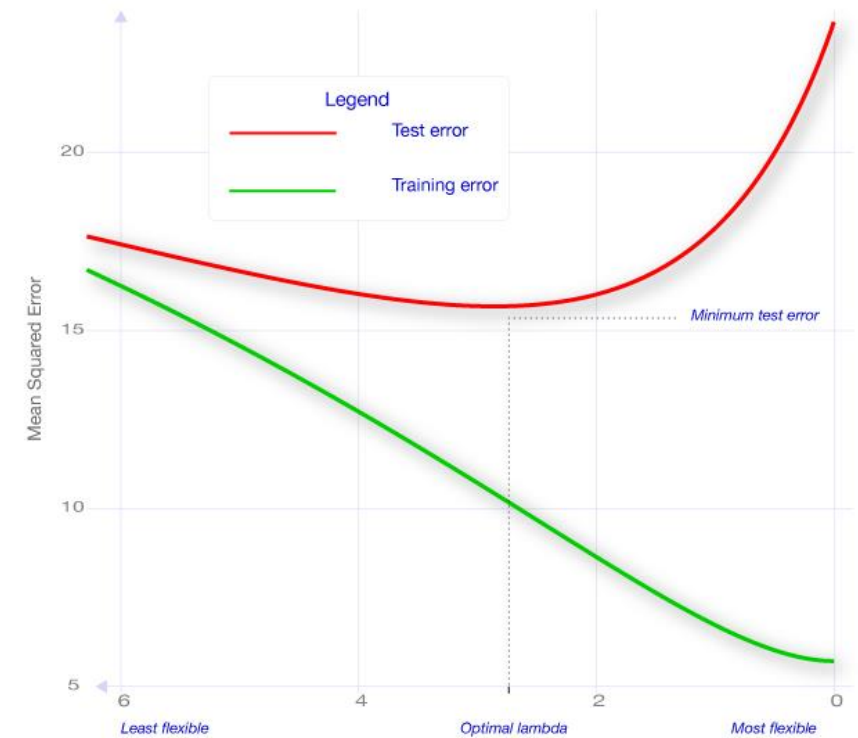
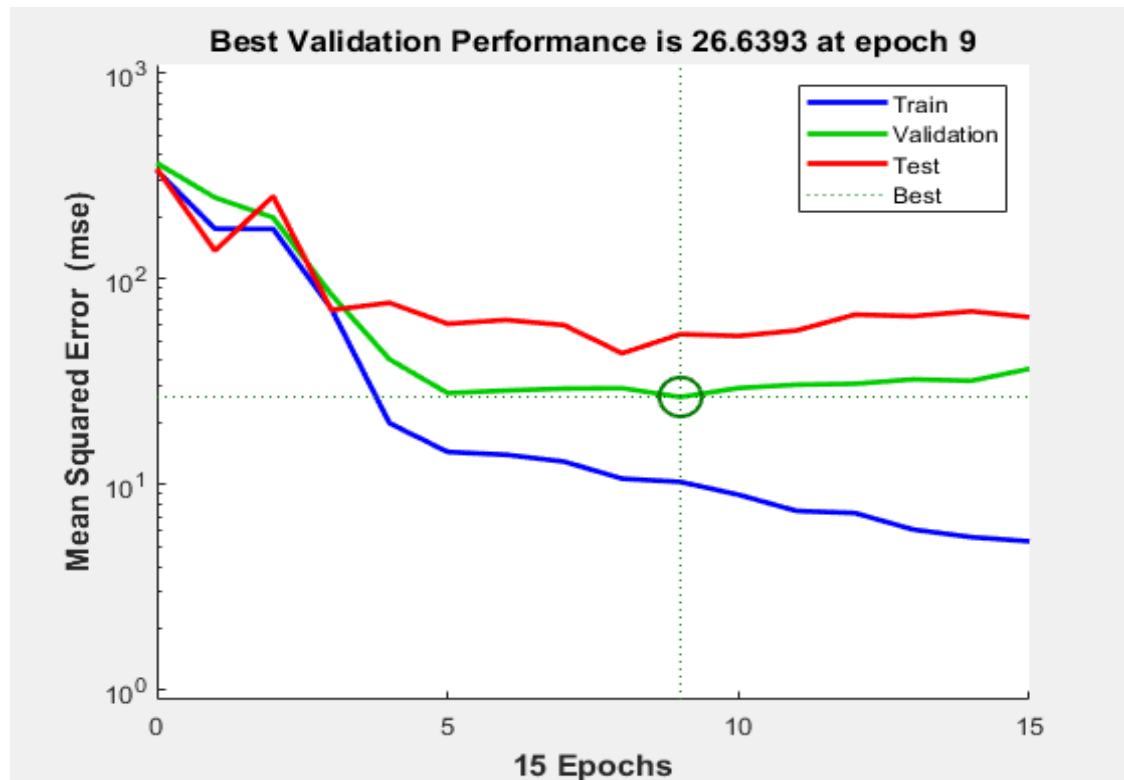
El aprendizaje o entrenamiento es el proceso en el cual los pesos sinápticos de la red son ajustados con el objetivo de capturar la información que se presenta, y de esta forma obtener respuestas adecuadas. Este proceso básicamente consiste en la presentación de un conjunto de datos, conocido como conjunto de patrones de entrenamiento, un número determinado de veces, conocido como ciclos, hasta que se produzca uno de los siguientes eventos:

- El error entre la salida de la red y la deseada alcance un valor aceptable.
- Se alcance el número máximo de ciclos.

El aprendizaje se lleva a cabo mediante el uso de algoritmos de entrenamientos. Los algoritmos de aprendizaje se pueden clasificar en dos tipos: supervisados y no supervisados.

Redes Neuronales Artificiales

- La forma óptima de ver el aprendizaje.



Redes Neuronales Artificiales

Algoritmos Supervisados

En los algoritmos supervisados, un conjunto de patrones de entrada y uno de patrones de salida son presentados a la red, de esta forma el conjunto de datos de entrenamiento está formado por el par (entradas, salida). El aprendizaje consiste en la modificación de los pesos sinápticos de manera de reducir la diferencia entre la salida de la red con la salida deseada. De acuerdo a la forma en que se ajustan los pesos sinápticos, este tipo de algoritmos se pueden clasificar en:

- **Algoritmo de aprendizaje por minimización de error:** se busca modificar los pesos de forma que se reduzca la diferencia entre la salida calculada por la red y la salida esperada. Se utiliza, por lo general, el error cuadrático medio como medida de la diferencia entre las salidas.
- **Algoritmos de aprendizaje por refuerzo:** busca minimizar el error al igual que el algoritmo anterior. Para esto, refuerza los pesos para resultados satisfactorios y penaliza los pesos para resultados malos. Este tipo de algoritmos se utiliza en casos en los que se dispone de información global sobre los patrones de entrenamiento, por ejemplo si son correctos o incorrectos.
- **Algoritmo de aprendizaje estocástico:** los pesos son modificados de acuerdo a cambios por lo general aleatorios, y luego es evaluado su efecto de acuerdo al resultado esperado.

Redes Neuronales Artificiales

Algoritmos no Supervisados

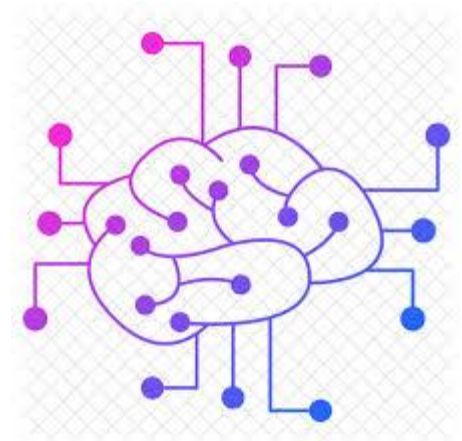
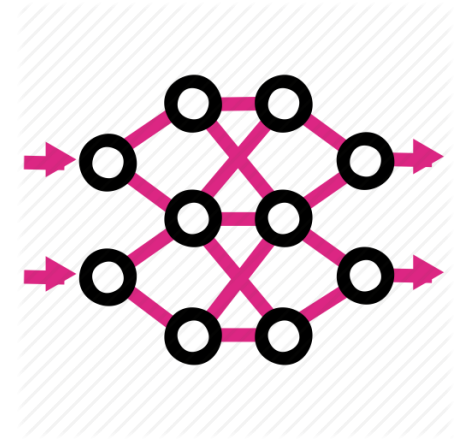
En los algoritmos no supervisados, un conjunto de patrones de entrada solamente es presentado a la red, formando el conjunto de datos de entrenamiento. De esta forma, el aprendizaje se realiza con base en los patrones de entrada, sin ser necesario indicar las salidas deseadas. Estos algoritmos de aprendizaje extraen ciertas propiedades de los patrones presentados y los agrupan en categorías de patrones similares. Los pesos sinápticos son modificados de manera tal que si se presentan dos patrones de datos similares se produzca la misma salida. En este tipo de algoritmos se pueden distinguir dos tipos de redes:

- **Redes de pesos fijos:** los pesos sinápticos son preestablecidos y pre calculados, por lo que no son adecuadas para utilizarlas en ambientes dinámicos. Las redes de Memoria Asociativa y las redes de Hopfield son ejemplos de este tipo de redes.
- **Redes de aprendizaje competitivos:** se realiza una competencia entre las neuronas y sólo se activan los pesos de la neurona ganadora. Las redes de auto-organización de Kohonen son un ejemplo de este tipo de redes.

Redes Neuronales Artificiales

3. Capacidad de Generalización

- La capacidad de generalización de una red tiene relación con la recuperación de la información que es almacenada en los pesos de las conexiones durante el entrenamiento, evaluando los resultados con un conjunto de datos diferente del utilizado en el proceso de aprendizaje, en el que pueden existir patrones diferentes.
- Se espera que cuando se presenten patrones que no han sido enseñados a la red, ésta sea capaz de entregar una respuesta cercana a la deseada. Mientras más precisa es la respuesta entregada por la red, más capacidad de generalización tendrá ésta.
- Para que una red neuronal sea capaz de generalizar de buena forma es necesario contar con un conjunto de datos de entrenamiento suficientemente representativo de la globalidad del problema en cuestión.



Redes Neuronales Artificiales

¿Qué pasa entonces en la aplicación de la RNA en series de tiempo?



Redes Neuronales Artificiales

La Red Neuronal Autorregresiva

En un modelo ARNN, la variable dependiente y_t es obtenida como una función no lineal de sus P valores pasados y_{t-p} , para $p = 1, \dots, P$:

$$y_t^* = \eta + \sum_{p=1}^P \varphi_p y_{t-p} + \sum_{h=1}^H \beta_h G(\omega_h + \sum_{p=1}^P \alpha_{p,h} y_{t-p})$$

donde $G()$ es la función sigmoidea adaptativa definida como:

$$G(u) = \left[\frac{1}{1 + \exp(-u)} \right]^M$$

Índice

1

Introducción

4

Neurona biológica vs
neurona artificial

2

Estadística, Machine
Learning, Deep Learning e
Inteligencia Artificial

5

Redes Neuronales
Artificiales

3

¿Por qué utilizar RNA en
series de tiempo?

6

Aplicación de las redes en
las series de tiempo

Aplicación de redes neuronales en las series de tiempo

Solemos usar la función `nnetar()` de la librería `forecast`.

```
nnetar(y, p, P = 1, size, repeats = 20, xreg = NULL, lambda = NULL, model = NULL, subset = NULL, scale.inputs = TRUE, x = y, ...)
```

y	A numeric vector or time series of class <code>ts</code> .
p	Embedding dimension for non-seasonal time series. Number of non-seasonal lags used as inputs. For non-seasonal time series, the default is the optimal number of lags (according to the AIC) for a linear AR(p) model. For seasonal time series, the same method is used but applied to seasonally adjusted data (from an <code>stl</code> decomposition).
P	Number of seasonal lags used as inputs.
size	Number of nodes in the hidden layer. Default is half of the number of input nodes (including external regressors, if given) plus 1.
repeats	Number of networks to fit with different random starting weights. These are then averaged when producing forecasts.

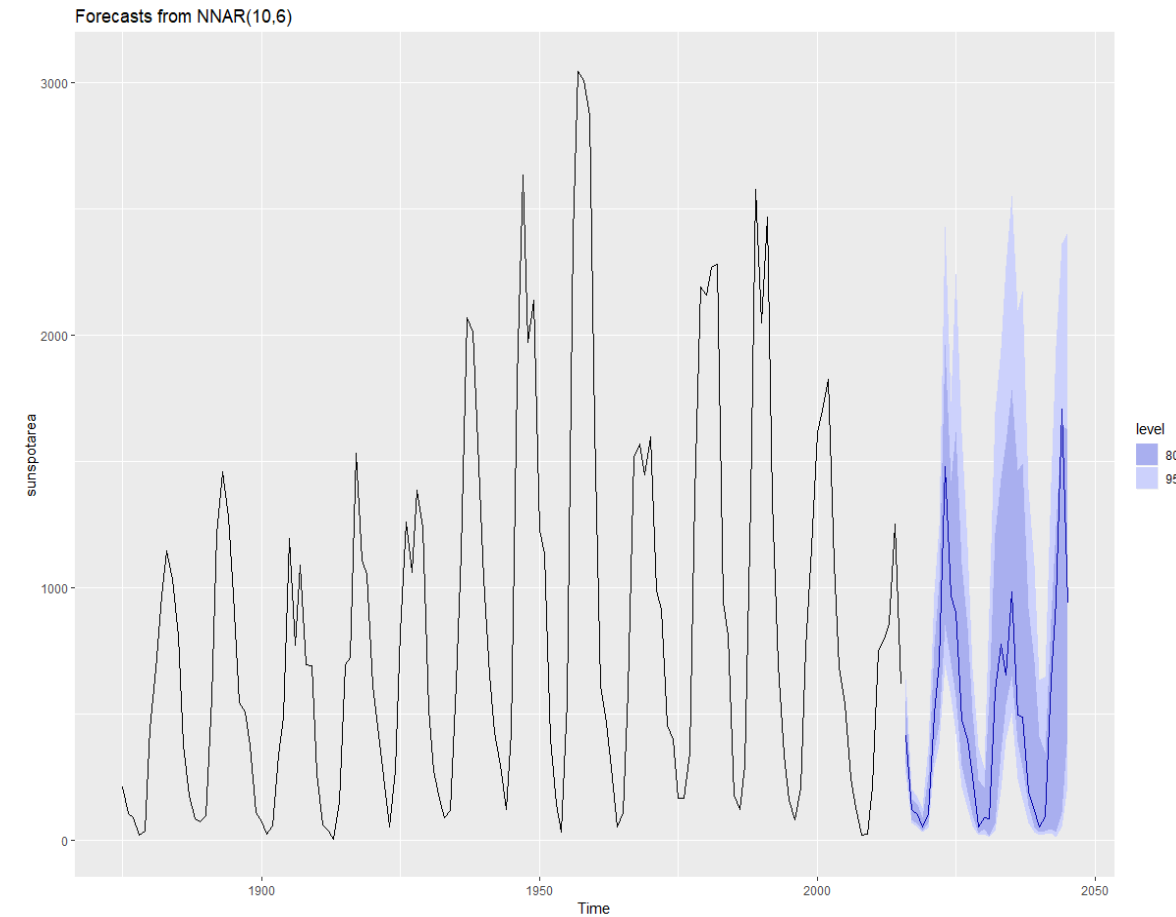
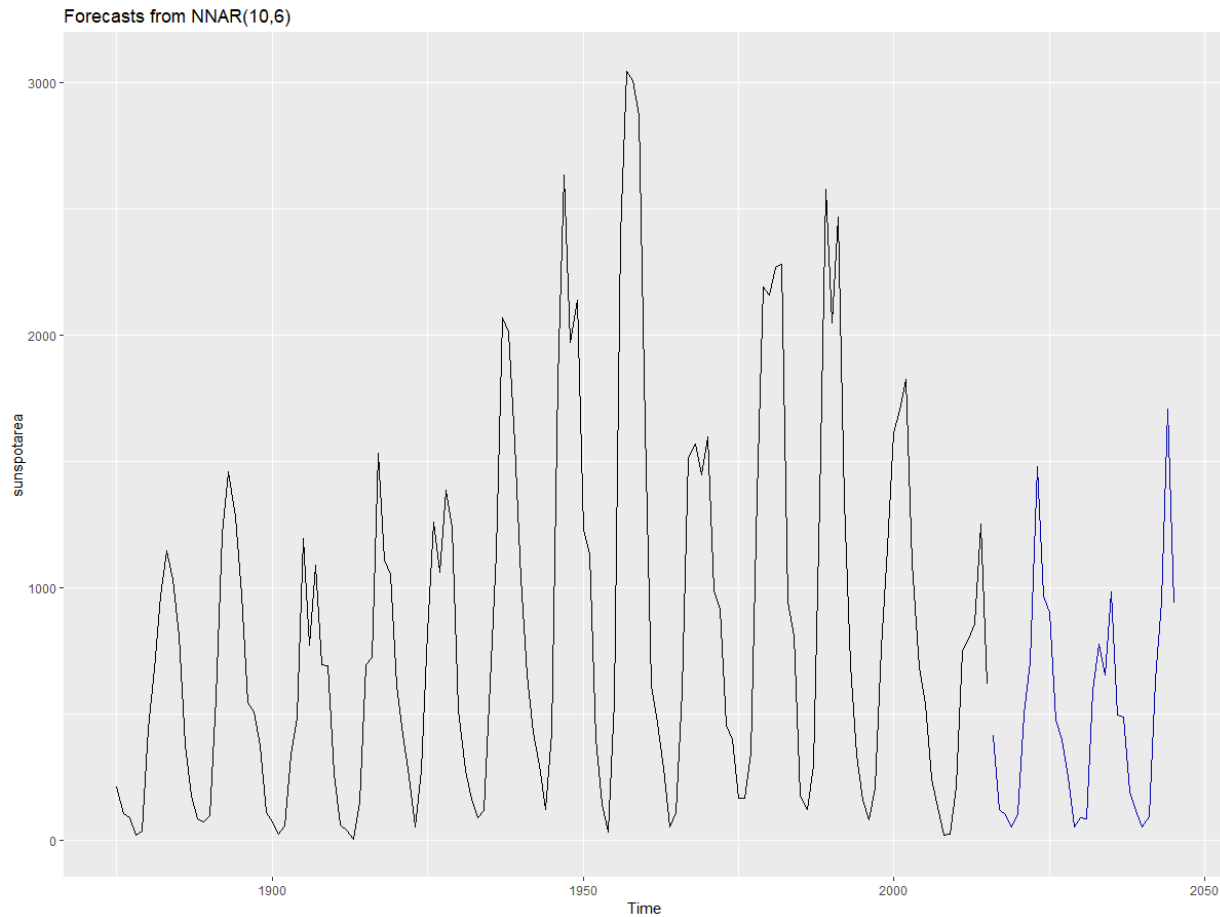
Ver los siguientes enlaces:

<https://www.rdocumentation.org/packages/forecast/versions/8.4/topics/nnetar>

<https://robjhyndman.com/hyndsight/nnetar-prediction-intervals/>

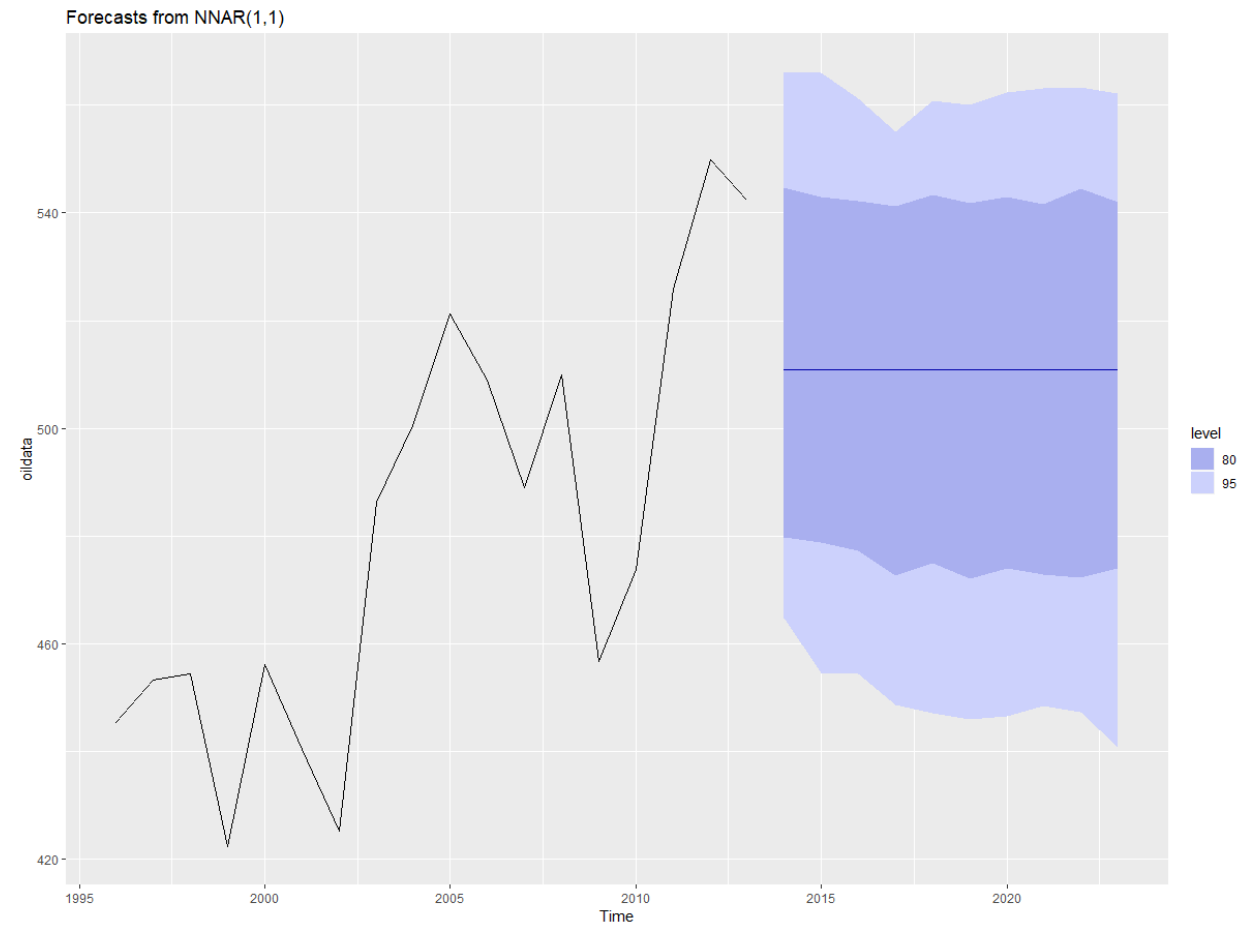
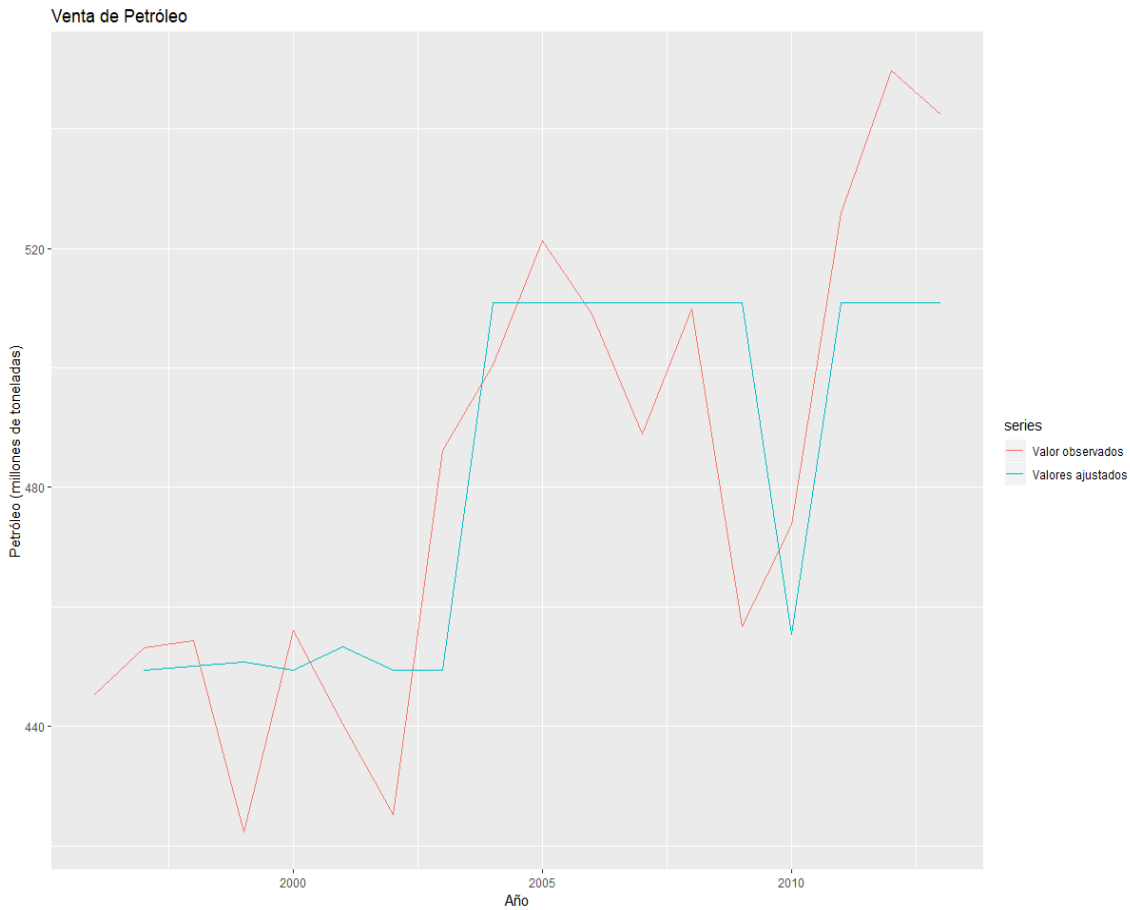
Aplicación de redes neuronales en las series de tiempo

Veamos la estimación y el pronóstico de la base de datos sunspotarea.



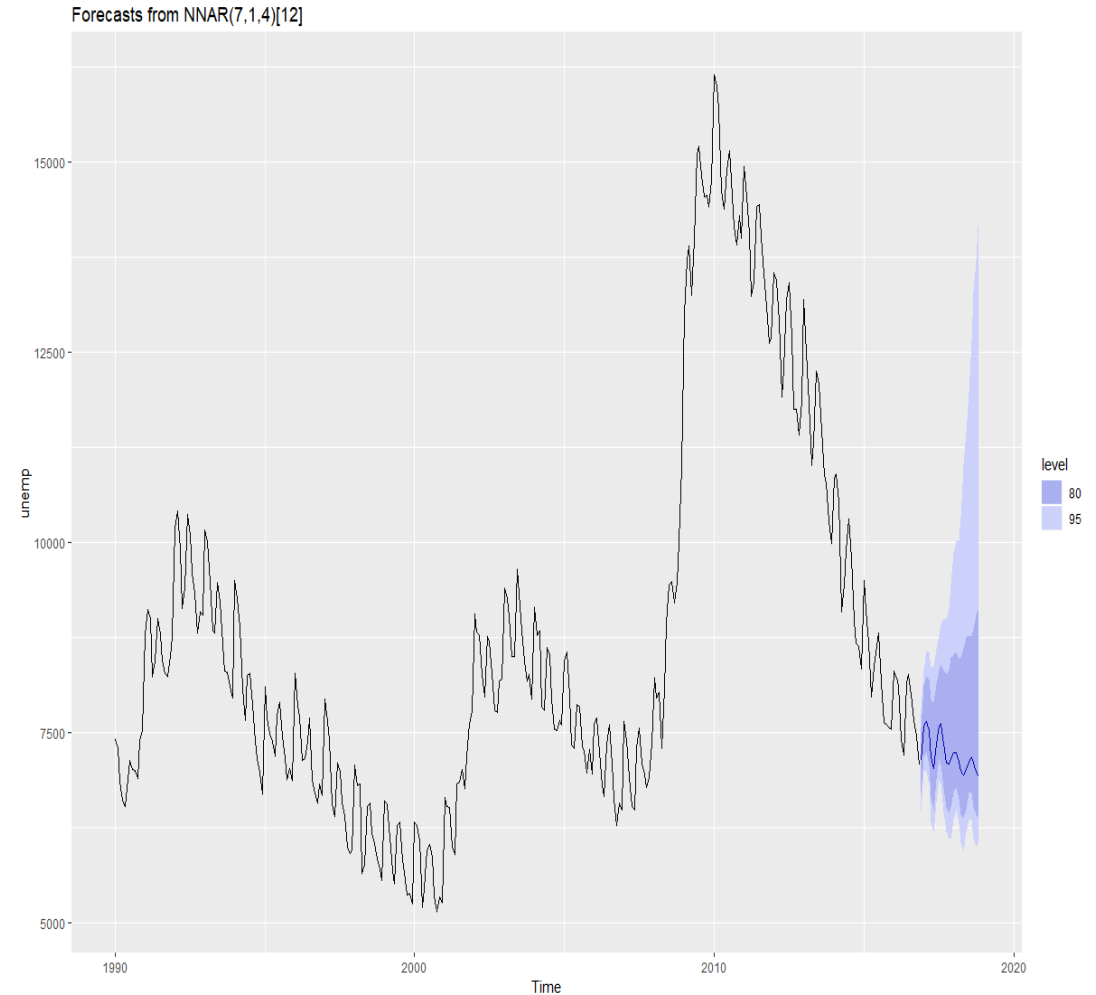
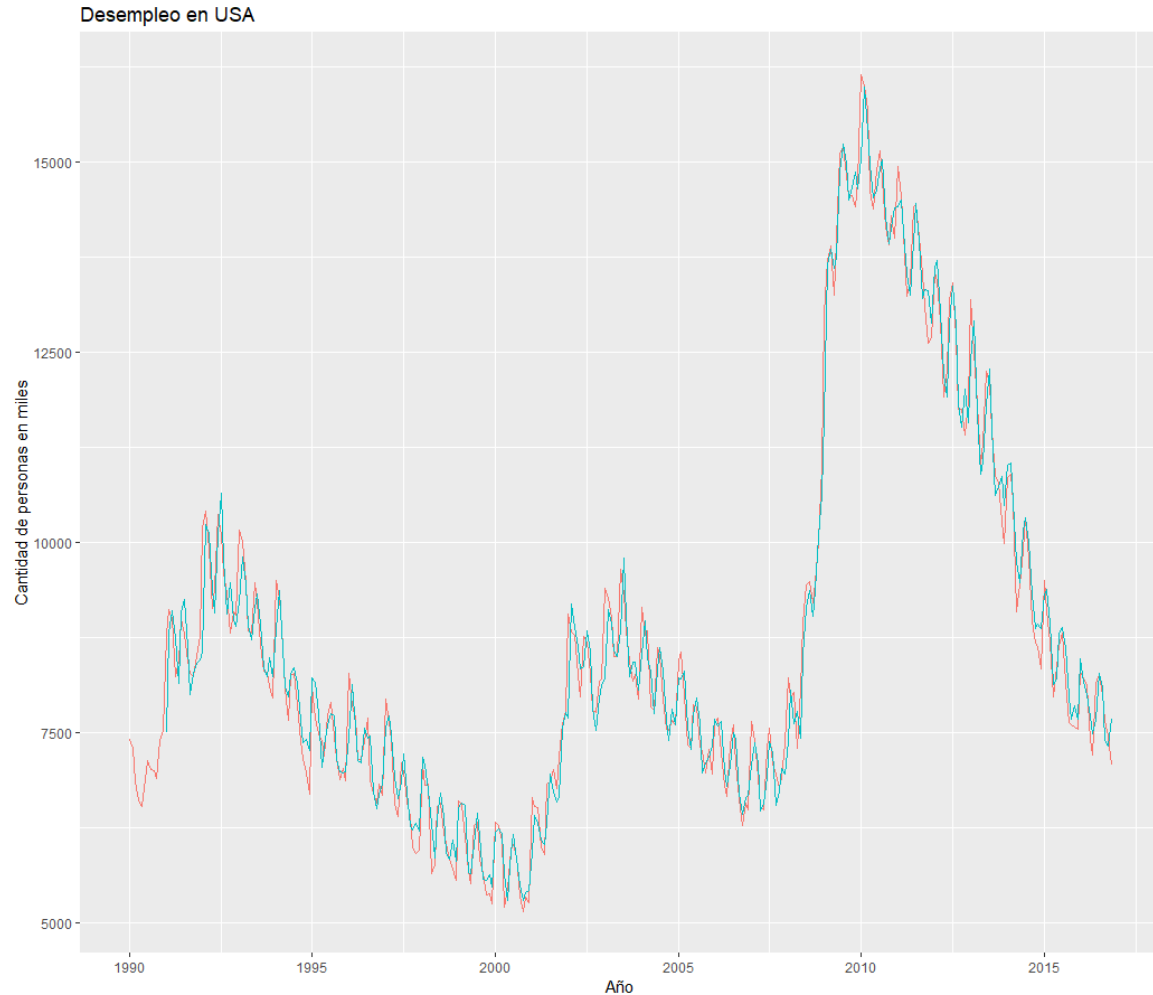
Aplicación de redes neuronales en las series de tiempo

Veamos la estimación y el pronóstico de la base de datos de oil. Para pocos datos las RNA no son nada efectivas...



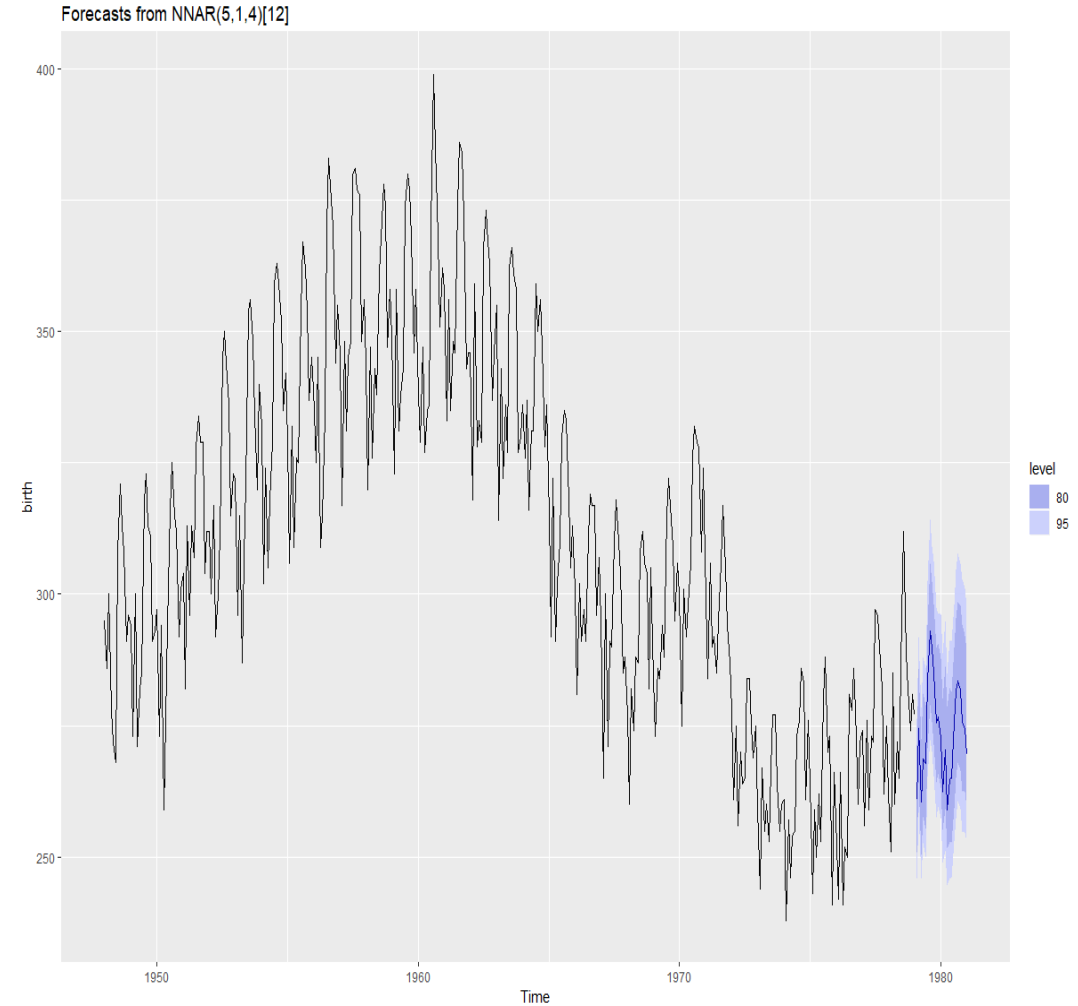
Aplicación de redes neuronales en las series de tiempo

Veamos la estimación y el pronóstico de la base de datos de unemp. No está tan mal, pero el ARIMA le gana...



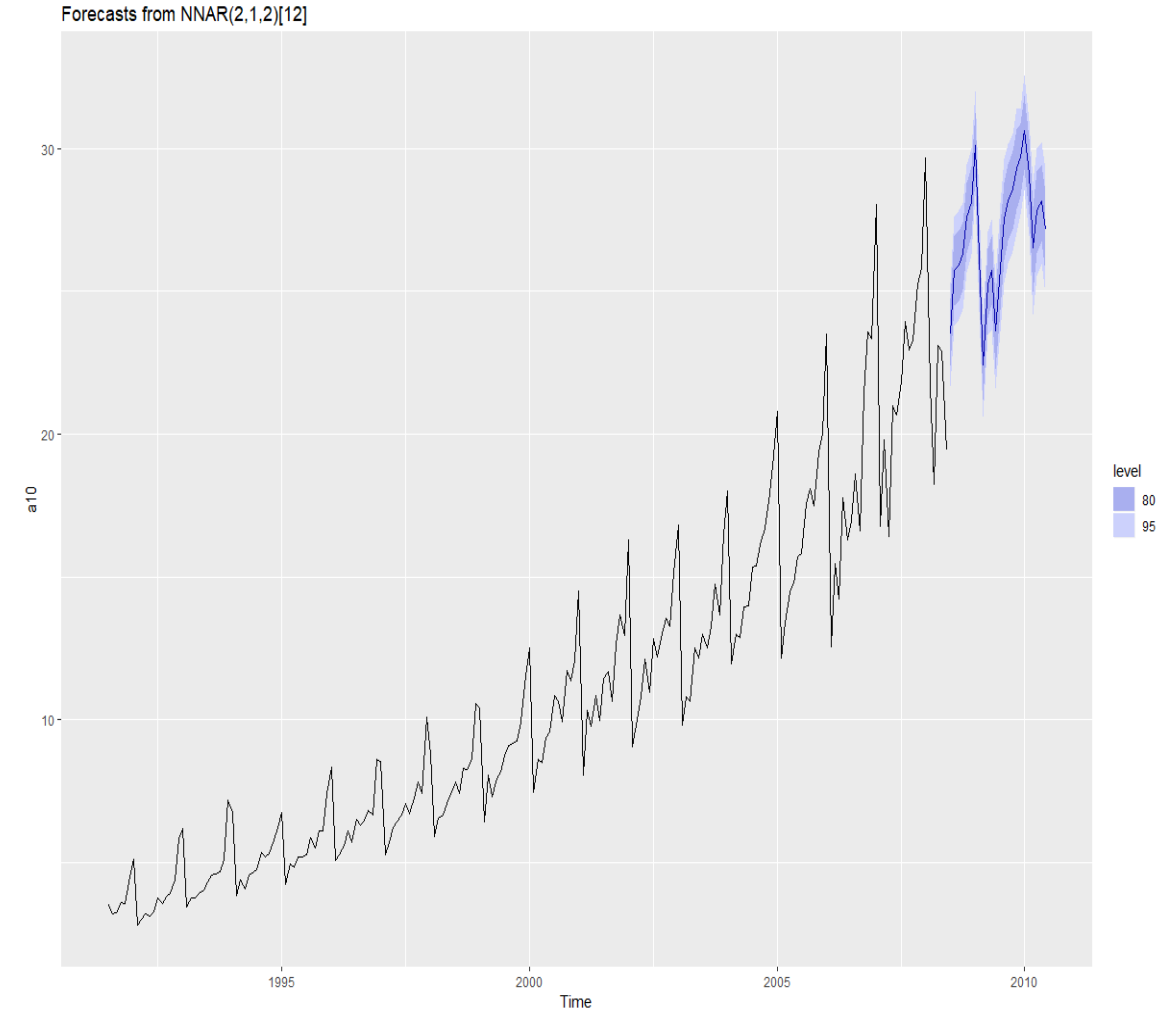
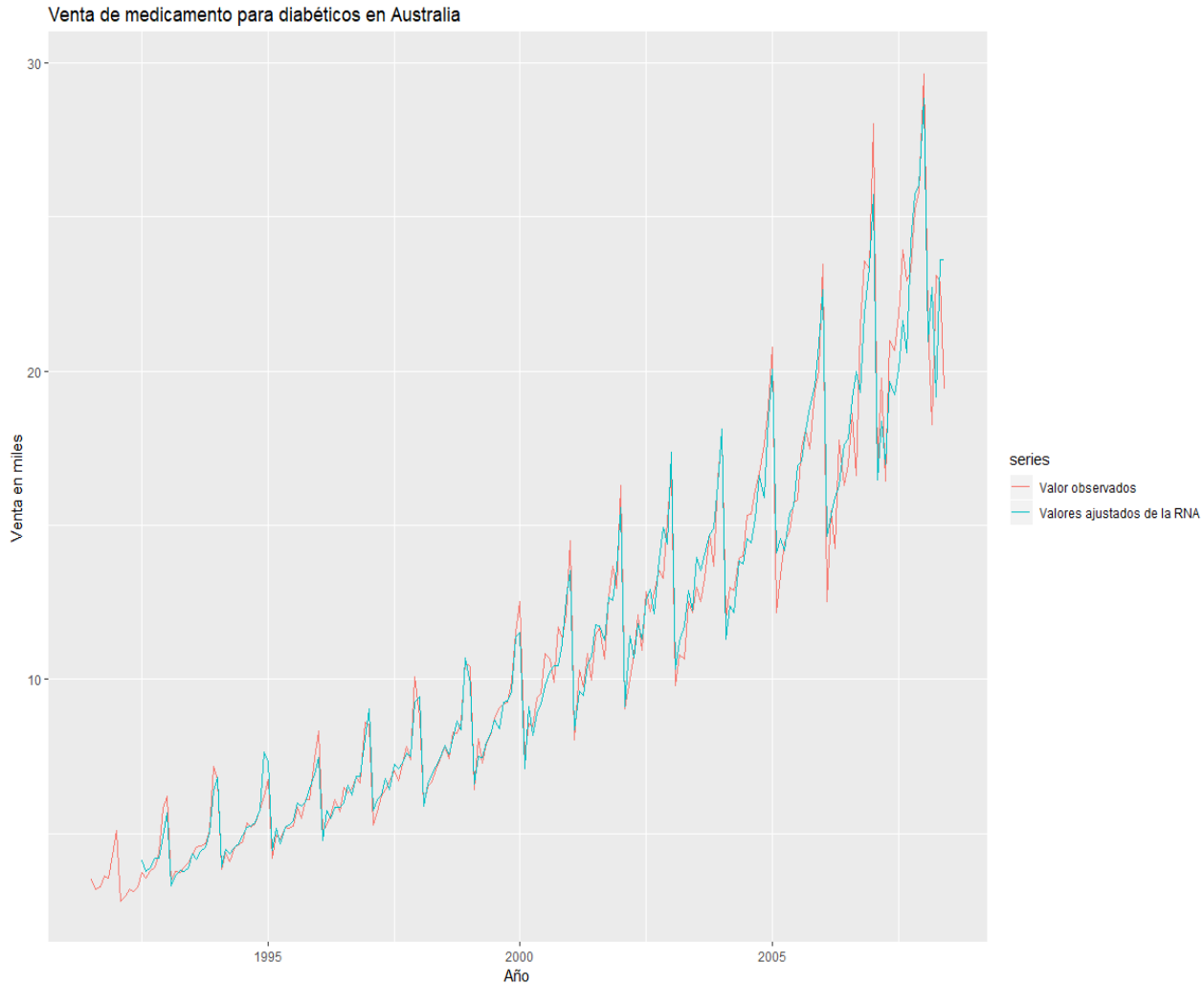
Aplicación de redes neuronales en las series de tiempo

Veamos la estimación y el pronóstico de la base de datos de birth. No está tan mal... pero el ARIMA le gana...



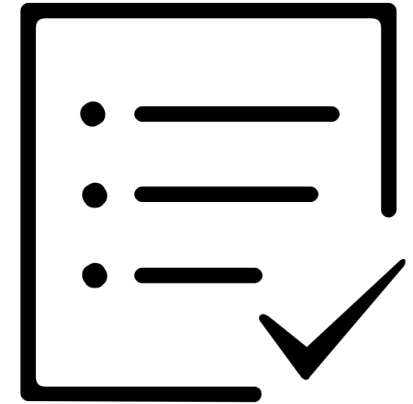
Aplicación de redes neuronales en las series de tiempo

Finalmente, veamos la estimación y el pronóstico de la base de datos de a10. Para mí hay un empate técnico...



Conclusión y discusión

- El presente tema trato de la aplicación de las redes neuronales en las series de tiempo.
- Se abordó la diferencia entre Estadística, Machine Learning, Deep Learning e Inteligencia Artificial.
- Las redes neuronales son otra alternativa en la aplicación del pronóstico en las series temporales.
- Este método se considera con mayor complejidad a los paramétricos, sin tener en cuenta el tipo de datos, la periodicidad.
- Algo que se debe de tener en cuenta es que las series pueden llegar a brindar un mejor resultado, sin embargo por su forma iterativa y de hallar las soluciones, requiere de mayor tamaño de períodos.



*The
End*