Introducción al Deep Learning

TÉCNICAS ESTADÍSTICAS PARA EL APRENDIZAJE II

Máster Universitario en Estadística Computacional
y Ciencia de Datos para la Toma de Decisiones

Introducción al Deep Learning

Introducción

Fundamentos biológicos

Arquitecturas de las redes neuronales artificiales

Historia de las redes neuronales artificiales

Limitaciones del *deep learning*

Introducción

¿Qué entendemos por **inteligencia**?
El Diccionario de la Real Academia define inteligencia como:
1. Capacidad de entender o comprender.
2. Capacidad de resolver problemas.
3. Conocimiento, comprensión, acto de entender.

¿Y por **Inteligencia Artificial**?
"Disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico".

Si diseñamos un algoritmo que sea capaz de encontrar una solución aceptable de forma rápida y casi siempre correcta, podríamos considerar que el **algoritmo** es **"inteligente"**.

Introducción

¿Qué tienen en común los **traductores automáticos** que nos permiten leer, en nuestra propia lengua, textos escritos en otros idiomas, los **sistemas de reconocimiento de voz** que todos llevamos en nuestro teléfono móvil o los **sistemas de visión artificial** que le permiten a un coche autónomo distinguir una señal de tráfico de otra?

Todos ellos son ejemplos de **aplicaciones** cotidianas basadas en el uso de **redes neuronales artificiales**.

Las redes neuronales artificiales, popularizadas actualmente bajo la denominación ***deep learning***, se enmarcan dentro del campo de la Inteligencia Artificial. Más concretamente, dentro de las técnicas de aprendizaje automático o *machine learning*.

Introducción

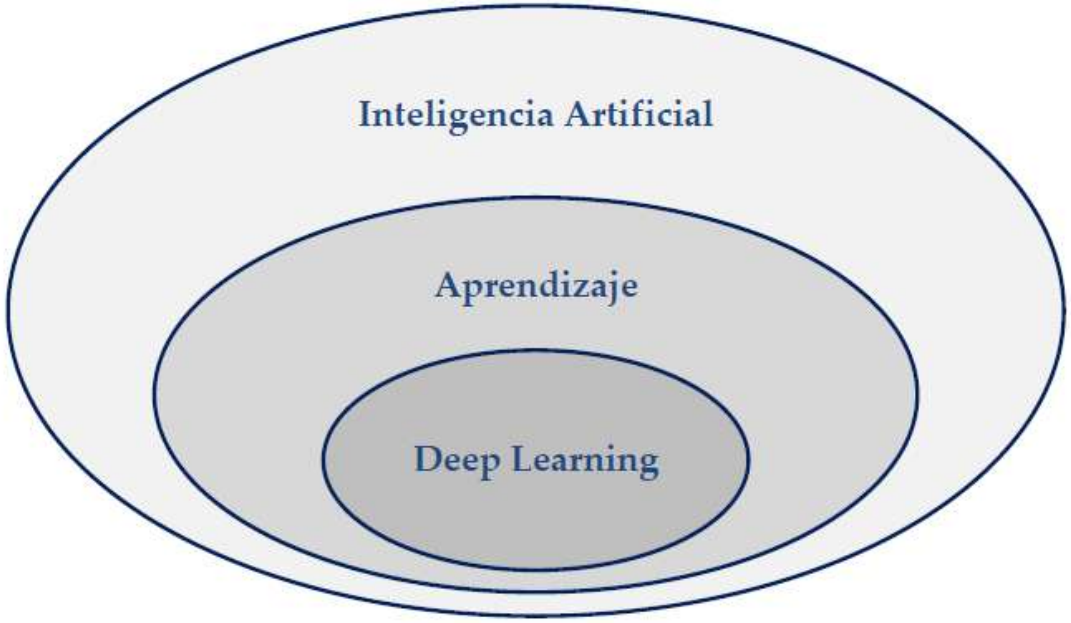
La **minería de datos** es un subconjunto de técnicas de aprendizaje automático que se pueden aplicar a conjuntos de datos enormes.

Las técnicas de aprendizaje se suelen combinar con métodos estadísticos y con técnicas de bases de datos que hagan posible el **análisis de cantidades ingentes de datos**, motivo por el cual, se suele hacer referencia a esta disciplina con el término genérico ***big data***.

La denominada **ciencia de datos** [***data science***] hace referencia al proceso de extracción de conocimiento denominado KDD, acrónimo de *Knowledge Discovery in Databases* (extracción no trivial de información potencialmente útil a partir de un gran volumen de datos, en el cual la información está implícita pero no se conoce previamente)

Los llamados **científicos de datos** disfrutan en la actualidad de una de las profesiones más prometedoras.

Introducción



Las técnicas *deep learning* están revolucionando el mundo de la Inteligencia Artificial y compañías como Google, Facebook, Amazon o Microsoft, entre otras muchas, se disputan a ingenieros, científicos de datos y doctorandos familiarizados con este tipo de técnicas.

Introducción

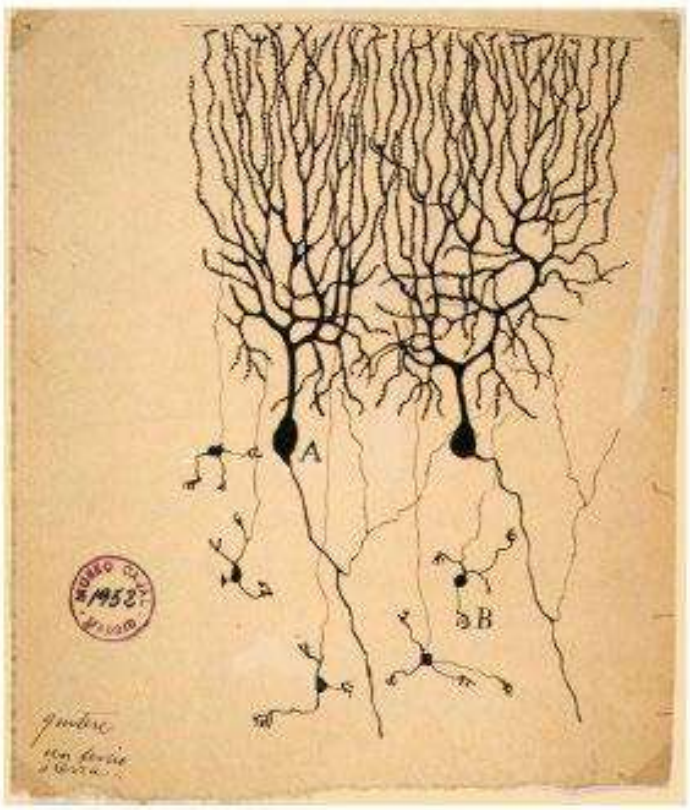
¿Qué es lo que diferencia al *deep learning* de otras familias de técnicas de aprendizaje?
Esencialmente, su **capacidad de abstracción**.

Los **algoritmos** de *deep learning* son de los más **robustos**. Se adaptan con facilidad a distintos tipos de datos, se entrenan con algoritmos relativamente sencillos y escalan bien a grandes conjuntos de datos.

Son, a día de hoy, una de las herramientas más versátiles de las que disponen los científicos de datos a la hora de **detectar patrones en conjuntos de datos**.

Las redes neuronales son capaces de **identificar y extraer las características que son realmente relevantes** y deben utilizarse para resolver un problema, algo que escapa de las posibilidades de otras técnicas de aprendizaje automático.

Fundamentos biológicos

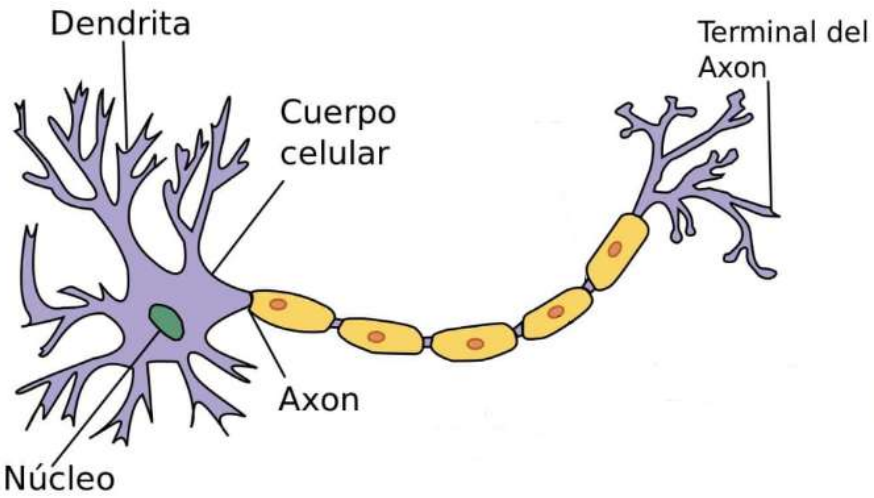


Neuronas del cerebelo
(Dibujo de Santiago Ramón y Cajal, 1899)

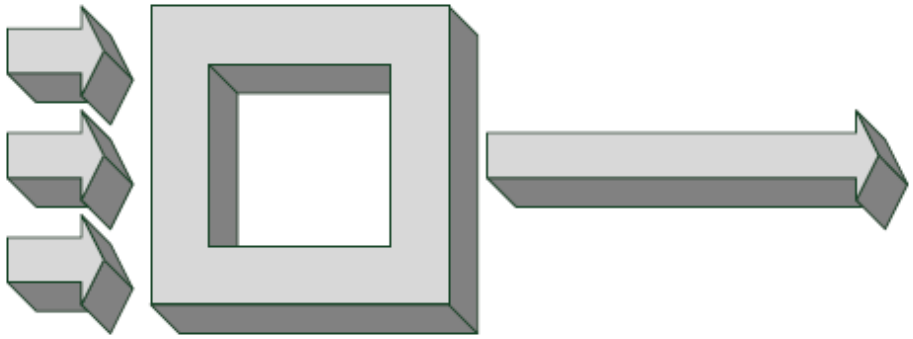
Durante décadas los científicos han perseguido la construcción de **algoritmos** capaces de procesar información al igual que el **cerebro humano**.

Las **redes neuronales artificiales** surgieron originalmente como una simulación abstracta de los sistemas nerviosos biológicos, constituidos por un conjunto de unidades llamadas neuronas conectadas unas con otras.

Fundamentos biológicos



Dibujo esquemático de las partes de una neurona biológica



Modelo abstracto de una neurona artificial

La **neurona** recibe una serie de señales de entrada a través de dendritas que la conectan a otras neuronas mediante sinapsis excitatorias e inhibitorias. El cuerpo o soma de la neurona combina e integra esas señales recibidas. En función de las circunstancias, la neurona es capaz de generar un pulso eléctrico [*spike*] de salida que se transmite a lo largo de su axón.

Fundamentos biológicos

El éxito de las **redes neuronales utilizadas en *deep learning*** se debe fundamentalmente a tres **características bioinspiradas**:

- Su **plasticidad**, que les permite adaptarse y aprender.
- Su **organización jerárquica**, que les faculta para resolver problemas complejos por composición.
- Su forma de modelar la **percepción**, que les proporciona mecanismos con los que resolver problemas de reconocimiento de patrones en señales de voz (reconocimiento de voz) y en imágenes visuales (identificación de objetos en visión artificial).

Fundamentos biológicos

Las redes neuronales artificiales se inspiran en lo (relativamente poco) que se sabe acerca del funcionamiento del cerebro humano.

Son capaces de resolver problemas que ningún programador humano había sido capaz de resolver con la ayuda de un ordenador.

Las redes neuronales artificiales están especialmente indicadas para **problemas complejos** en los que existen multitud de casos particulares.

Se aplican en numerosas áreas como:

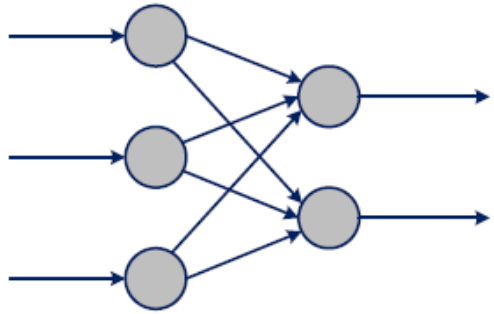
- Reconocimiento de patrones (imagen, voz, caracteres)
- Robótica
- Medicina
- Predicción de series temporales …

Fundamentos biológicos

Desde el punto de vista computacional, puede que lo único que nos interese es que **la señal proveniente de una entrada afecta al estado de una neurona**.

Amplificada o atenuada de diferentes formas, se combina con otras señales de entrada para determinar el **nivel de activación** de la neurona. En función de ese nivel de activación, la salida de la neurona podrá variar (o no).

De ahí que las sinapsis se modelen habitualmente como un simple número real, conocido normalmente como **peso** asociado a la conexión.



El resultado: una red formada por **nodos que representan neuronas y enlaces que representan sinapsis**. Los enlaces, con pesos, dan lugar a un grafo ponderado, generalmente dirigido.

Ejemplos:

Visión artificial

Reconocer un objeto tridimensional dentro de una escena, con condiciones de iluminación cambiantes.

Detección de fraudes

Calcular la probabilidad con la que una transacción realizada con una tarjeta de crédito es fraudulenta.

No existen reglas simples que sean fiables (hay que combinar múltiples reglas que no siempre indican la presencia de fraude). Además, los tipos de fraude van cambiando, por lo que el programa que los detecte debe ir evolucionando.

Vehículos autónomos

...

Solución

En vez de diseñar un algoritmo que resuelva el problema, **recopilamos** un montón de **datos** (ejemplos).

Diseñamos un **algoritmo que aprenda de esos datos** y cree el programa necesario para resolver el problema.

El programa generado automáticamente no tiene por qué parecerse a un programa implementado manualmente (en el caso de las redes neuronales, puede contener millones de números reales).

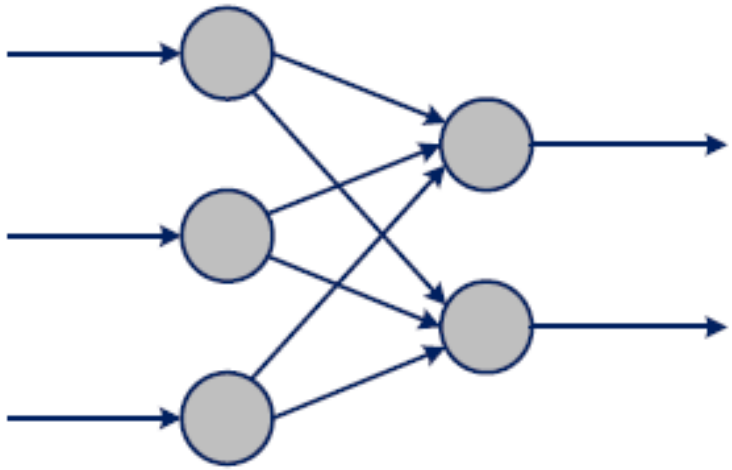
Si tenemos éxito, el programa funcionará bien para nuevos ejemplos, aunque sean diferentes a los que utilizamos para su entrenamiento.

Si los datos cambian, el **programa puede cambiar** entrenándolo de nuevo.

Arquitecturas de las redes neuronales artificiales

Existe un gran número de **arquitecturas diferentes** de redes de neuronas artificiales.

La **topología más habitual** de las redes neuronales está formada por **múltiples capas**, cada una de las cuales recibe como entrada la salida de la capa anterior. De esa forma, se consigue un sistema modular en el que cada capa proporciona un mayor nivel de abstracción con respecto a la entrada que recibe la red neuronal.



Arquitecturas de las redes neuronales artificiales

Por **ejemplo**, cuando se trabaja con imágenes:

La entrada de la red será una matriz de píxeles, cada uno de ellos con un nivel de intensidad asociado. En el caso de las imágenes en color, la red recibirá un conjunto de matrices, una por cada canal de la imagen (rojo, verde, azul).

En una primera capa, la red será capaz de identificar dónde aparecen bordes en la imagen, fronteras entre zonas con diferente nivel de intensidad en sus píxeles.

Capas posteriores serán capaces de combinar fronteras para formar esquinas, posteriormente formar objetos, escenas...

Y así hasta que la red sea capaz de identificar la presencia de un objeto concreto.

Arquitecturas de las redes neuronales artificiales

Receta de un algoritmo de *deep learning*:

- Recopilar un conjunto de **datos** asociado al problema (enorme, si es posible).
- Diseñar una **función** de coste apropiada para el problema, también conocida como función de pérdida [*loss function*].
- Seleccionar un **modelo** de red neuronal y establecer sus hiperparámetros (tamaño, características...).
- Aplicar un algoritmo de **optimización** para minimizar la función de coste ajustando los parámetros de la red.

Historia de las redes neuronales artificiales

1943 Primer modelo de neurona artificial. Neurona de McCulloch-Pitts

1957 Primer algoritmo de aprendizaje supervisado (Perceptrón)

1969 Análisis de las capacidades y limitaciones del perceptrón. La investigación en redes neuronales casi desaparece

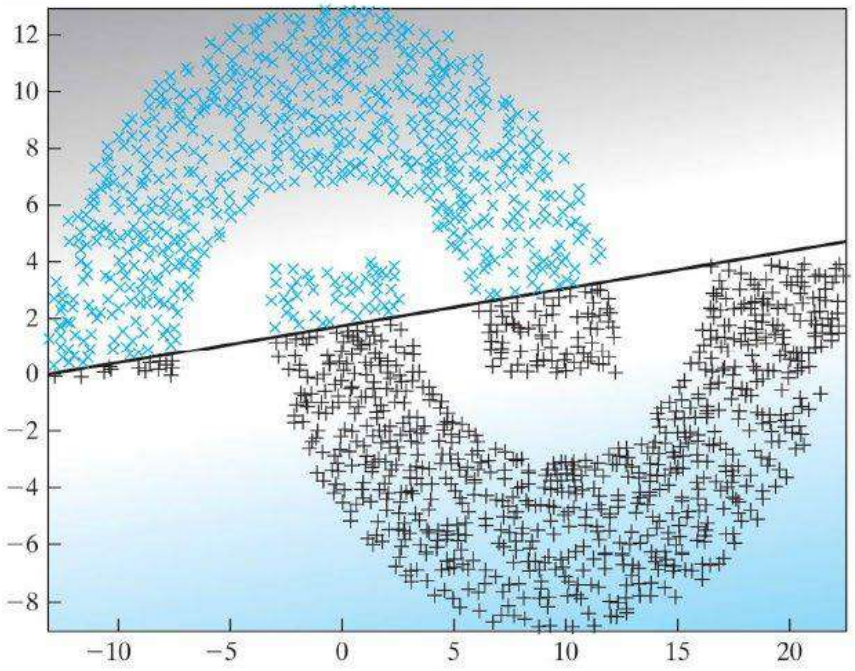
1982 Redes recurrentes (tratan datos secuenciales de forma eficiente; tienen *memoria*)

1986 Se populariza el uso del algoritmo *backpropagation*. Algoritmo de entrenamiento de redes multicapa. Renacimiento de las redes neuronales artificiales

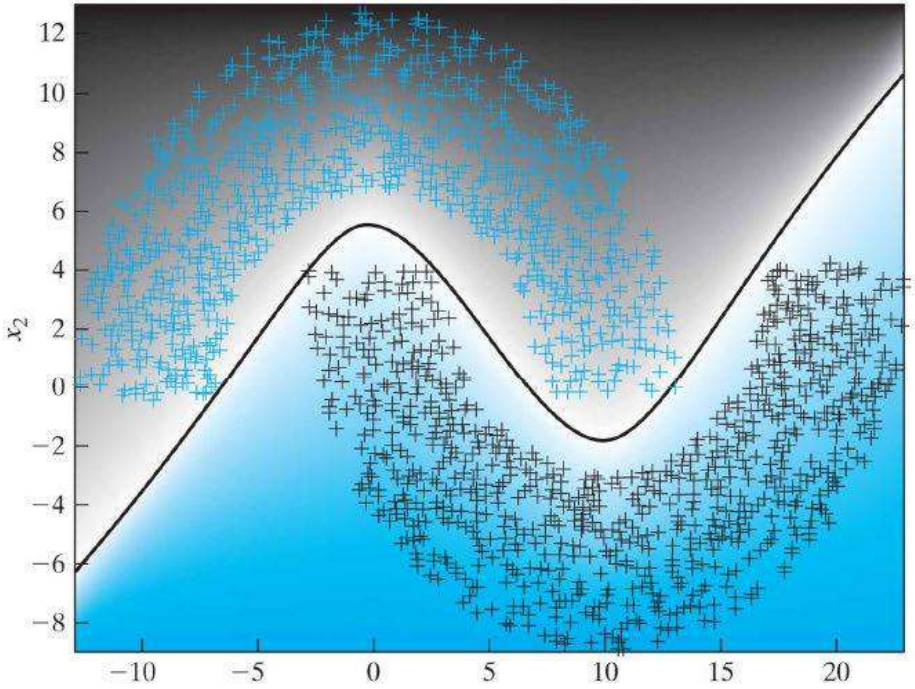
1990 Redes convolutivas (visión artificial, reconocimiento de objetos en imágenes)

2006 *Deep Learning*. Se desarrollan formas de entrenar redes mucho más grandes con decenas de capas de neuronas artificiales

Historia de las redes neuronales artificiales



Perceptrón



Red multicapa

Historia de las redes neuronales artificiales

No fue hasta la primera década del siglo XXI cuando se popularizaron los modelos de redes neuronales artificiales con **múltiples capas ocultas**, denominadas "redes neuronales profundas" [*deep neural networks*], las que dieron lugar al *deep learning*.

Aunque los orígenes de las redes neuronales artificiales se remontan a los comienzos de la Inteligencia Artificial, en los años 40 del siglo XX, y el *deep learning* en sí despegaría en 2006, el éxito real del *deep learning* se debe a su **aplicación práctica en problemas de interés para la industria tecnológica**, como el reconocimiento de voz o la visión artificial.

En la última década, se ha conseguido igualar, cuando no mejorar, el rendimiento de los seres humanos en la resolución de tareas que, hasta hace poco, se consideraban de nuestro dominio exclusivo.

Historia de las redes neuronales artificiales

En esencia, las redes neuronales utilizadas en la actualidad no difieren tanto de las que se usaban hace 30 años.

Entonces, ¿por qué se ha popularizado tanto el *deep learning*? **¿Qué factores han contribuido a revitalizar las redes neuronales artificiales?**

Tres son las causas principales:

- Disponibilidad de conjuntos de datos enormes
- Potencia de cálculo proporcionada por las GPU [*Graphical Processing Unit*]
- Desarrollo de nuevos algoritmos

Y, ¿qué **ventaja** ofrecen las redes neuronales profundas?
Básicamente, la posibilidad de representar la misma función que una red neuronal poco profunda, pero hacerlo de forma más económica, con menos unidades ocultas.

Limitaciones del *deep learning*

Sus dos aspectos más problemáticos son los derivados del **sobreajuste** o sobreaprendizaje y su carácter de **cajas negras** (incapacidad para determinar cómo llegan las redes neuronales a una conclusión).

Para resolver el primero, que es un problema común con otros muchos modelos de aprendizaje automático, se han propuesto multitud de técnicas que, más o menos, nos permiten soslayarlo.

En cuanto al segundo, es algo sobre lo que todavía queda mucho por hacer y que puede tener implicaciones con respecto a la seguridad de los sistemas que emplean redes neuronales internamente.

Limitaciones del *deep learning*

El **sobreaprendizaje** se produce siempre que un modelo se ajusta tan bien a su conjunto de entrenamiento que **deja de generalizar correctamente** cuando lo utilizamos sobre un conjunto de prueba diferente. Es un problema habitual en muchas técnicas de aprendizaje automático.

Cuando utilizamos **redes neuronales**, éstas incluyen **multitud de parámetros ajustables**: los pesos que modelan las conexiones entre neuronas. Ese número elevado de parámetros las hace propensas a sufrir problemas de sobreaprendizaje [*overfitting*]. De hecho, hay quien piensa que es su principal inconveniente.

Como veremos, existe una amplia gama de técnicas y herramientas a nuestra disposición para **prevenir el sobreaprendizaje** al que son propensas las redes neuronales.

Limitaciones del *deep learning*

En cuanto a su carácter de **cajas negras:**

En entornos altamente regulados, como las **finanzas**, la ley puede obligar a una compañía a que explique las razones tras una decisión tomada con ayuda de un modelo automatizado. ¿Por qué se le deniega la concesión de un crédito a un posible cliente? ¿Por qué se le incrementa el importe de la póliza de un seguro médico?

En **aplicaciones militares de defensa o de seguridad**, este hecho puede suponer un hándicap, especialmente a posteriori si se producen accidentes y se busca al responsable.
¿Por qué se pilota un vehículo autónomo de esa forma y no de otra?
¿Por qué se señala a alguien particular como sospechoso de terrorismo si, en principio, parece llevar una vida completamente normal?
¿Es un caso flagrante de discriminación por edad, sexo, raza o religión?

Probablemente, surjan nuevas técnicas que lo que hagan sea, una vez tomada una decisión, inventar una explicación plausible que sirva para cubrir el expediente legal.

En resumen…

Las Redes Neuronales Artificiales:

- No se programan, se entrenan.
- Necesitan disponer de ejemplos, en un número suficiente y una distribución representativa para ser capaces de generalizar correctamente.
- Requieren de un proceso de validación para evaluar la "calidad" del aprendizaje conseguido.

Veremos cómo:

- Entrenar redes neuronales artificiales (para distintos modelos de red).
- Preparar (preprocesar) los ejemplos necesarios para su entrenamiento.
- Evaluar la calidad del proceso de aprendizaje.

Bibliografía

•[1] Fernando Berzal. Redes Neuronales & Deep Learning. Edición independiente.

•[2] François Chollet. Deep learning with Python. Manning Shelter Island.

•[3] Ian Goodfellow, Yoshua Bengio & Aaron Courville. Deep Learning . MIT Press.