

Machine Learning II

¿Qué es Deep Learning?

U-TAD

Contexto Deep Learning

¿En qué punto nos encontramos hoy?

Es un subdominio del Machine Learning, que a su vez es un subdominio de la IA.

El Deep Learning se caracteriza porque la mayoría de los algoritmos que forman parte de este subdominio se corresponden con redes neuronales artificiales profundas.

Este dominio es probablemente a día de hoy el dominio que más se está utilizando dentro de la IA para resolver problemas reales.

Tenemos algunos problemas que se están resolviendo con redes neuronales y con deep learning como, por ejemplo, clasificación de imágenes, conducción autónoma, clasificación de texto.

Hay muchísimos problemas que estamos resolviendo a día de hoy con Deep Learning y probablemente sea el subdominio de la IA que mayor comunidad investigativa tiene y donde más recursos se están invirtiendo.

Contexto Deep Learning

Nosotros nos vamos a centrar en los fundamentos del deep learning.

Es decir, qué es una neurona artificial, qué es una red neuronal artificial, cómo son los componentes y cómo funcionan matemáticamente. Como funcionan de manera formal y, por supuesto, vamos a ver de manera práctica cómo implementarlos para resolver problemas reales.

Por tanto, nos quedaremos con la arquitectura fundamental cuando hablamos de deep learning, que son las redes neuronales densamente conectadas. No obstante, hay otros tipos de redes neuronales (como las convolucionales y recurrentes) que, posiblemente, no debieran empezar a estudiarse hasta no tener muy clara toda la base que rodea a las redes neuronales artificiales profundas.

Por lo tanto, si nos planteamos la pregunta de qué tipo de problemas puedo resolver utilizando técnicas de deep learning, la respuesta es que una gama muy amplia, desde problemas simples, como identificar el objeto que aparece en una imagen, hasta problemas complejos como la simulación de vídeos o de fotos.

¿Qué es el Deep Learning?

Primero, repasemos algunos conceptos de machine learning e IA:

- Desde los inicios de la IA, esta tecnología rápidamente abordó y resolvió problemas que son intelectualmente difíciles para los seres humanos pero relativamente sencillos para las computadoras.
- El verdadero reto para la IA demostró ser resolver las tareas que son fáciles de realizar para las personas pero difíciles de describir formalmente.

Un ejemplo podría ser el ajedrez, que aunque es un juego complicado para los seres humanos, porque requiere el cálculo de diferentes variaciones a la hora de realizar una jugada, lo podemos describir en base a un conjunto formal de reglas.

Es decir, cada una de las piezas se mueve de una manera determinada. Tenemos un número de casillas determinado, tenemos unas piezas determinadas y, por tanto, es una tarea que un ordenador con un algoritmo basado en IA o ML puede resolverla de manera relativamente sencilla. Y, sin embargo, para los seres humanos es algo muy complejo de resolver.

¿Qué es el Deep Learning?

Sin embargo, hay otras tareas como podría ser, por ejemplo, identificar un objeto que aparece en una imagen, que para los seres humanos es algo prácticamente intuitivo (nosotros vemos una imagen y podemos identificar lo que aparece en ella).

Pero, por el contrario, es muy difícil de describir formalmente a través de un conjunto de reglas fijas, cuando aparece un objeto determinado en esa imagen o cuando aparece otro objeto diferente.

Resultando que estos tipos de problemas, los segundos, son los que realmente suponen un reto para los algoritmos basados en IA.

¿Qué es el Deep Learning?

- La disciplina más popular dentro de la IA es el **Machine Learning**.

Y, concretamente, los algoritmos basados en ML resuelven un problema utilizando datos de experiencia pasada de cómo se resolvió ese problema en particular.

- El rendimiento de estos algoritmos de aprendizaje automático depende en gran medida de la representación de los datos que se les proporciona.

Ejemplo: si utilizamos un algoritmo de ML para diagnosticar a un paciente, el sistema de IA no examina al paciente directamente, en su lugar, el médico le indica al sistema varios datos relevantes, como la presencia, o ausencia, de determinados síntomas de un paciente. Y el algoritmo, en base a esas características de entrada que se le proporcionan, tomará una decisión de si ese paciente está enfermo o no lo está.

Y para ello, se le debió proporcionar previamente información o experiencia pasada de otros pacientes que enfermaron y de cuáles eran los síntomas que tenían en el pasado.

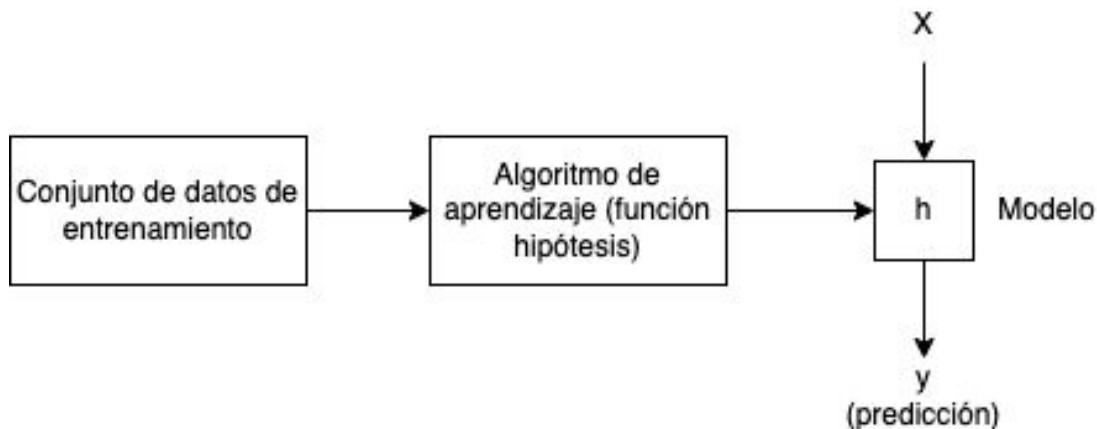
¿Qué es el Deep Learning?

Definición más formal del Machine Learning:

“El aprendizaje automático es un subdominio de la IA que proporciona a los sistemas la capacidad de aprender y mejorar automáticamente a partir de la experiencia **sin ser explícitamente programados para ello**. Se basa en la hipótesis subyacente de **crear un modelo** y tratar de **mejorarlo ajustando más datos en el modelo** a lo largo del tiempo.”

¿Qué es el Deep Learning?

En qué consiste crear un modelo:



Como ya vimos, este es el funcionamiento genérico de una técnica de ML y también será, aunque con algunas sutilezas, el funcionamiento de las técnicas de deep learning.

A continuación veremos las diferencias entre las técnicas de ML que vimos en el 1Q y las de deep learning que veremos en este 2Q.

End-to-End Learning

Para explicar las diferencias entre el ML y el DL, comenzaremos por el concepto de E2E learning: surge como consecuencia de un problema que tenemos con las características de entrada.

- **El rendimiento de estos algoritmos de ML depende en gran medida de la representación de los datos que se les proporciona.**

Como sabemos, la extracción de características es fundamental para que nuestro algoritmo realice buenas predicciones. Y esa extracción previa de las características depende de un analista (tiene que decidir qué características son relevantes).

- Muchas tareas pueden resolverse diseñando un conjunto adecuado de características a extraer para esa tarea y proporcionándolas a un algoritmo de ML.
- Sin embargo, **para otras tareas, es difícil identificar qué características deben ser extraídas.**

End-to-End Learning

Por ejemplo, supongamos que quisiéramos escribir un programa para detectar coches en fotografías. Sabemos que los coches tienen retrovisores, por lo que nos gustaría utilizar la presencia de un retrovisor como una característica (o focos, ventanas, limpiaparabrisas, etc)

Desafortunadamente, es difícil describir exactamente cómo se ve un retrovisor en términos de valores de píxeles. Su imagen puede variar por la posición, el color, por las sombras o el sol al deslumbrar, u oculto por otro objeto...)



End-to-End Learning

Es decir, hay determinados problemas al fin y al cabo, como por ejemplo la detección de objetos en una imagen, que si tratamos de extraer características para identificar ese objeto, es una tarea realmente complicada.

Y existen algoritmos de ML para los que tenemos que extraer estas características que probablemente no funcionarían bien.

End-to-End Learning

- Una solución a este problema es utilizar técnicas de ML para descubrir la mejor forma de describir los datos. Este enfoque se conoce como **Representation Learning**.

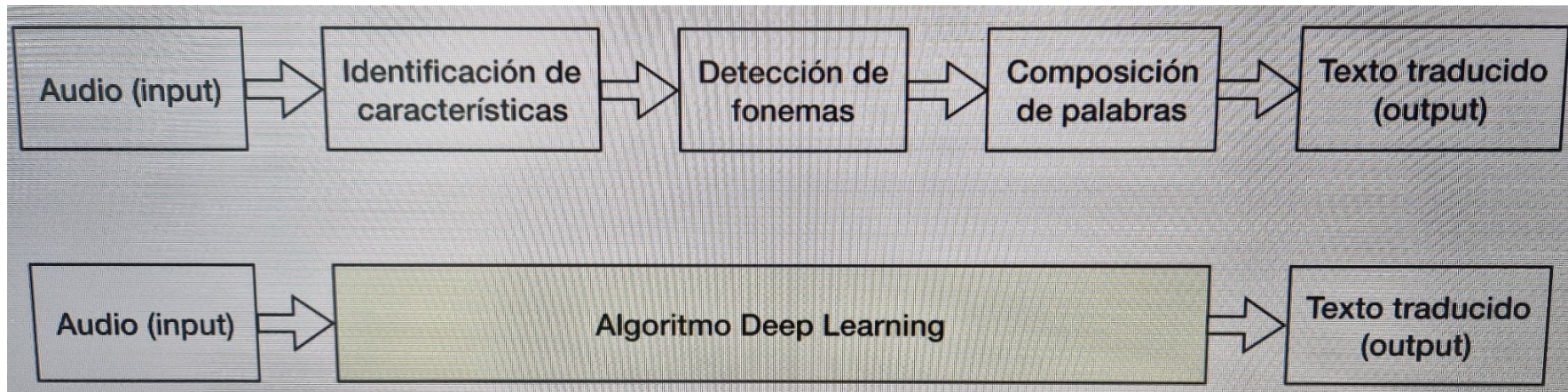
Van a coger la imagen y van a proporcionar una serie de características que muchas veces para los humanos no son comprensibles, pero que de alguna manera con ese algoritmo digamos que tiene cierto sentido y que representan bien esa información.

- El ejemplo por excelencia de un algoritmo basado en **Representation Learning** es el algoritmo **Autoencoder** (lo veremos más adelante).

A pesar de estas técnicas, en determinadas ocasiones, sigue siendo muy difícil extraer características de alto nivel y abstractas de los datos en bruto. Determinados factores, como el acento de un orador (si queremos obtener su país de procedencia), pueden ser identificados únicamente teniendo una comprensión sofisticada, casi a nivel humano, de los datos (idioma).

- Cuando es casi tan difícil obtener una representación que resolver el problema original, las técnicas de **Representation Learning** no son de utilidad.

End-to-End Learning



¿Cómo lo resolvemos? Con la técnica denominada **End-to-End Learning**

Con las técnicas clásicas de ML, tomamos el audio, identificamos una serie de características de ese audio, se las proporcionan a un algoritmo de ML con la que se detectan fonemas. Después se realiza una composición de palabras y, por último, se traduce el texto.

Con E2E Learning, y concretamente este tipo de técnica es la que van a utilizar la mayoría de los algoritmos basados en Deep Learning, se va a recibir el audio en bruto, proporcionándose al algoritmo de DL, y este algoritmo va a ser capaz de darnos como output el texto directamente traducido.

Toda esas fases intermedias va a realizarlas el propio algoritmo. Y esto es E2E Learning.

Caso práctico: Intuiciones sobre E2E Learning

resources/Intuiciones sobre E2E Learning.ipynb

(pip install tensorflow) // No es necesario entender el notebook

Antes de continuar con la teoría, vamos a dedicar unos minutos a comprender de manera un poquito más práctica esas intuiciones que hemos visto en los apartados anteriores. Concretamente, el concepto de **E2E learning** y cómo podemos resolver un problema que mediante la creación de reglas o de extracción de características de manera manual puede complicarse bastante y con técnicas de deep learning, aplicando este concepto de E2E Learning, podemos llegar a una solución bastante aceptable en un tiempo muy reducido.

El conjunto de datos consiste en una serie de imágenes de 28x28 píxeles en las que aparece una prenda de ropa.

El problema que queremos resolver consiste en clasificar una de las diez categorías de prendas que aparecen.

Caso práctico: Intuiciones sobre E2E Learning

2. Visualización del conjunto de datos

Si tratásemos de resolver este problema utilizando reglas estáticas, con if-elses, sería demasiado complicado porque tendríamos que ir analizando la intensidad de los píxeles. También sería complejo con las técnicas tradicionales de ML si tenemos que hacer una extracción de características. ¿Cuáles extraeríamos?

Sería bastante complicado encontrar características que, a priori, puedan ser significativas, para determinar si en la imagen aparece una prenda u otra.

Precisamente para esto se aplica el concepto de **E2E Learning**

Tendremos un subconjunto de entrenamiento que tiene 60k imágenes que están clasificadas (etiquetadas) como pertenecientes a una de esas diez clases en función de la prenda que aparece en ellas. Y otro subconjunto de pruebas, con el que realizaremos la validación de nuestra red neuronal, de 10k imágenes.

No es necesario entender lo siguiente en detalle porque iremos viendo los conceptos más adelante.

Caso práctico: Intuiciones sobre E2E Learning

3. Transformación del conjunto de datos

Con técnicas de DL, simplemente, aplicaremos una serie de transformaciones de nuestros datos, que en este caso se corresponden con coger nuestras imágenes y transformarlas en un único vector que tendrá una longitud de 28×28 .

Se normaliza, es decir, se sitúa en una escala similar (normalmente entre 0 y 1), y se transforman esas etiquetas de acuerdo a ese valor que nos dice si aparece una camiseta, una zapatilla, ..., recibiendo un valor del 0 al 9.

Esto se codifica de una manera diferente, en forma de vector.

Es algo rápido y estándar para la mayoría de los problemas de Deep Learning.

Caso práctico: Intuiciones sobre E2E Learning

4.- Construcción del modelo

Ahora vamos a entrenar una red neuronal artificial que está formada, en este caso, por cuatro capas: una primera capa formada por 300 neuronas, una segunda de 100 neuronas y una tercera de 10. Y, de alguna manera, la primera capa que es la input layer (que veremos más adelante), y que son las características de entrada, en este caso, el valor de cada uno de los píxeles.

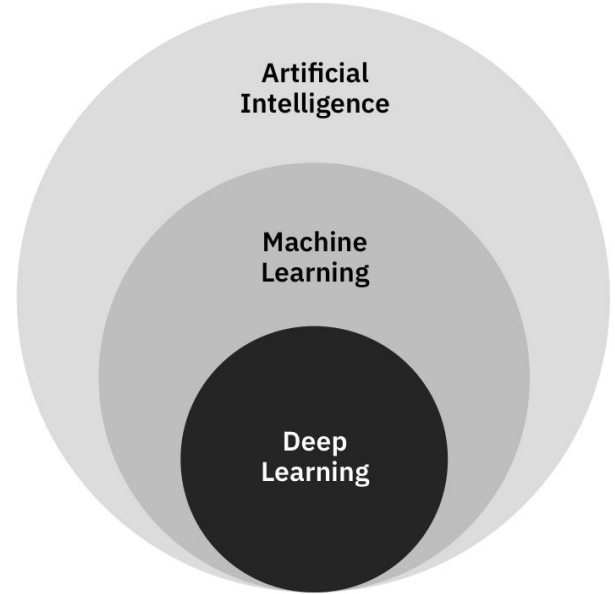
Lo importante es que una vez se define la red neuronal y se va entrenando, lo que me va a decir por un lado es la precisión y la exactitud que tiene a la hora de clasificar estas imágenes en las diferentes clases en función de lo que aparece, indicándonos cuanto ha acertado.

Para el último entrenamiento, tenemos una exactitud de un 89% y una precisión del 90% para ejemplos que no se han utilizado en el entrenamiento.

En resumen, hemos conseguido resolver aceptablemente este problema de clasificación de imágenes en unos 5 minutos. ¿Cuánto implicaría analizando los píxeles?

Cont: ¿Qué es el Deep Learning?

El Deep Learning es un subdominio del Machine Learning que resuelve el problema de las técnicas de **Representation Learning** introduciendo representaciones que se expresan en términos de otras representaciones más simples (lo veremos en la siguientes slides). Se fundamenta en la construcción de conceptos complejos a partir de conceptos más simples.



¿Qué es el Deep Learning?

¿Esto quiere decir que nos podemos olvidar de todo aquello de construcción de un modelo, función hipótesis y demás?

No, de hecho, las técnicas de DL van a funcionar en este aspecto de manera muy similar a las técnicas de ML, con la diferencia de que cuando nosotros teníamos el conjunto de datos de experiencia pasada para las técnicas de ML, aquí hacíamos una extracción, una selección, de características de entrada.

Una predicción muy concienzuda que en general marcaba muchísimo la predicción o la efectividad de nuestro algoritmo de ML.

En este caso, en nuestro conjunto de datos de entrenamiento, vamos a tener información o datos en bruto que vamos a proporcionarle a nuestro algoritmo DL, y ese algoritmo será el que se encargue de seleccionar las características que considere más oportunas para resolver el problema, y de proporcionar al final el resultado.

¿Qué es el Deep Learning?

Sin embargo, seguiremos teniendo parámetros del modelo.

El algoritmo seguirá correspondiendo con una serie de funciones matemáticas que veremos muy en detalle a lo largo de este cuatrimestre, y ajustaremos los parámetros de estas funciones en este proceso, que hemos denominado como entrenamiento, para llegar a construir un modelo para ajustar nuestra función h .

Y, una vez que tenemos esa función ajustada o ese modelo, podremos coger información nueva, un ejemplo nuevo, una instancia nueva y realizar una predicción.

Con lo cual, como podemos observar, realmente los pasos generales son muy similares. Lo que cambia, en esencia, es que si antes nosotros queríamos predecir si aparecía un objeto o no en una imagen y empezábamos a extraer características de ese objeto, como la intensidad de los píxeles que representan una parte del objeto o algo así, en este caso, no vamos a hacer nada de eso: vamos a utilizar toda la imagen, proporcionándosela al algoritmo, y este decidirá qué es importante y qué no para diferenciar entre las imágenes.

¿Qué es el Deep Learning?

De manera resumida, si volvemos al ejemplo que veíamos un coche, cuando nosotros utilizábamos una técnica de ML, teníamos que coger esta imagen, realizar la extracción de características de alguna forma (identificando un foco, una rueda, la luna, etc). Después proporcionamos esas características a nuestro algoritmo de ML y este nos decía si aparece un coche, o no, en esa imagen.

Input -> Feature extraction -> Classification -> Output

Con las técnicas de DL cambia, porque cogemos la imagen, le indicamos en esa imagen si aparece un coche, o no, para ese proceso de entrenamiento. Le proporcionamos esa imagen a nuestro algoritmo de DL, y este se encarga de realizar todo el proceso hasta que nos devuelve una predicción indicando si en esa imagen aparece un coche o no.

Input -> Feature extraction + Classification -> Output

¿Qué es el Deep Learning?

¿Esto quiere decir que las técnicas de DL, como, por ejemplo, las redes neuronales profundas, no pueden utilizarse en un problema en el que nosotros hagamos una extracción de características porque pensemos que pueda ser útil?

Por ejemplo, pensamos que para la predicción de si una transacción bancaria es fraudulenta, o no, en lugar de pasarle toda esa predicción a un algoritmo, pensamos que haciendo una primera extracción de características podría ser más útil y puede dar mejores resultados (incluso en un algoritmo de DL).

¿Podemos utilizar un algoritmo de DL de esta forma?

Sí, también podemos utilizar un algoritmo de DL realizando esa fase de extracción de características. Sin embargo, en muchas ocasiones, si podemos resolver un problema realizando extracción de características con un algoritmo de DL, también podemos resolverlo con una técnica más sencilla de ML que, probablemente, consuma menos recursos y nos proporcione resultados muy similares.

¿Qué es el Deep Learning?

Con lo cual, ¿para qué vamos a dedicar estos algoritmos de Deep Learning?

Vamos a dedicarlos y orientarlos más a problemas para los que realizar esta primera extracción de características sea algo realmente complejo. Como, por ejemplo, la detección de imágenes, la detección de texto, la identificación de lo que se está diciendo en una audio, etc.

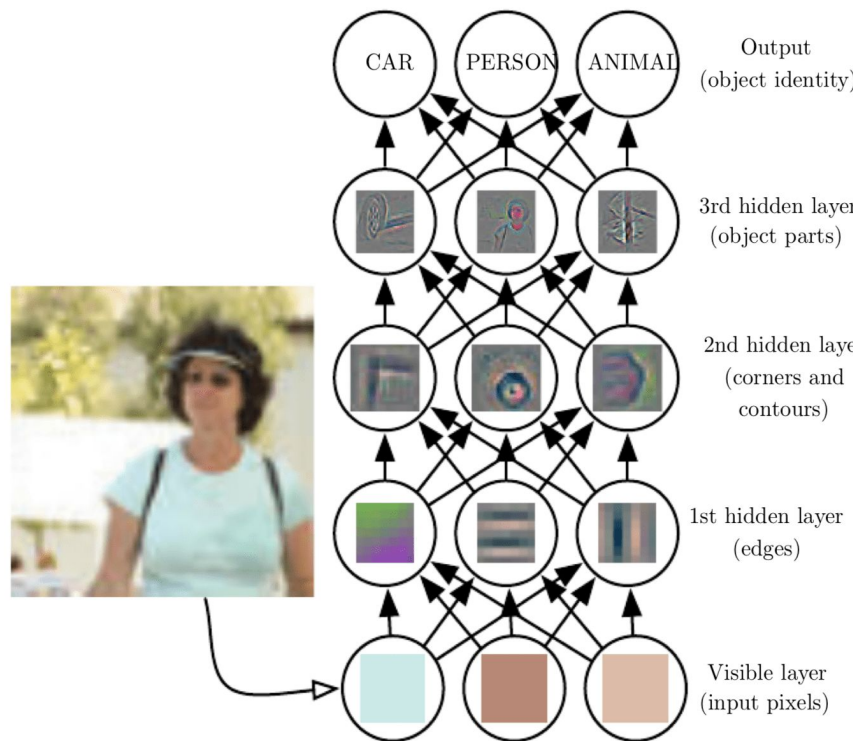
¿Qué es el Deep Learning?

Esta imagen muestra de alguna manera el significado de lo que hemos comentado anteriormente.

El DL se basa en redes neuronales: vamos a coger datos, por ejemplo, una imagen en la que aparece una persona, y vamos a proporcionársela a una red neuronal que tendrá diferentes capas y diferentes neuronas artificiales por cada una de las capas (desde el input hasta el output layer)

Estas redes neuronales van a ir aprendiendo, extrayendo características, de esa imagen. Y, a medida que va avanzando en las capas de la red neuronal, se van a ir componiendo características más complejas.

La primera se encargará de extraer, por ejemplo, colores. La siguiente identifica texturas, combinaciones de colores. En la siguiente, aparecen contornos. Y, a medida que avanzamos, aparecen partes humanas (oreja, cara, ...). Y, finalmente, en la última capa se decide si en la imagen aparece un coche, una persona o un animal.

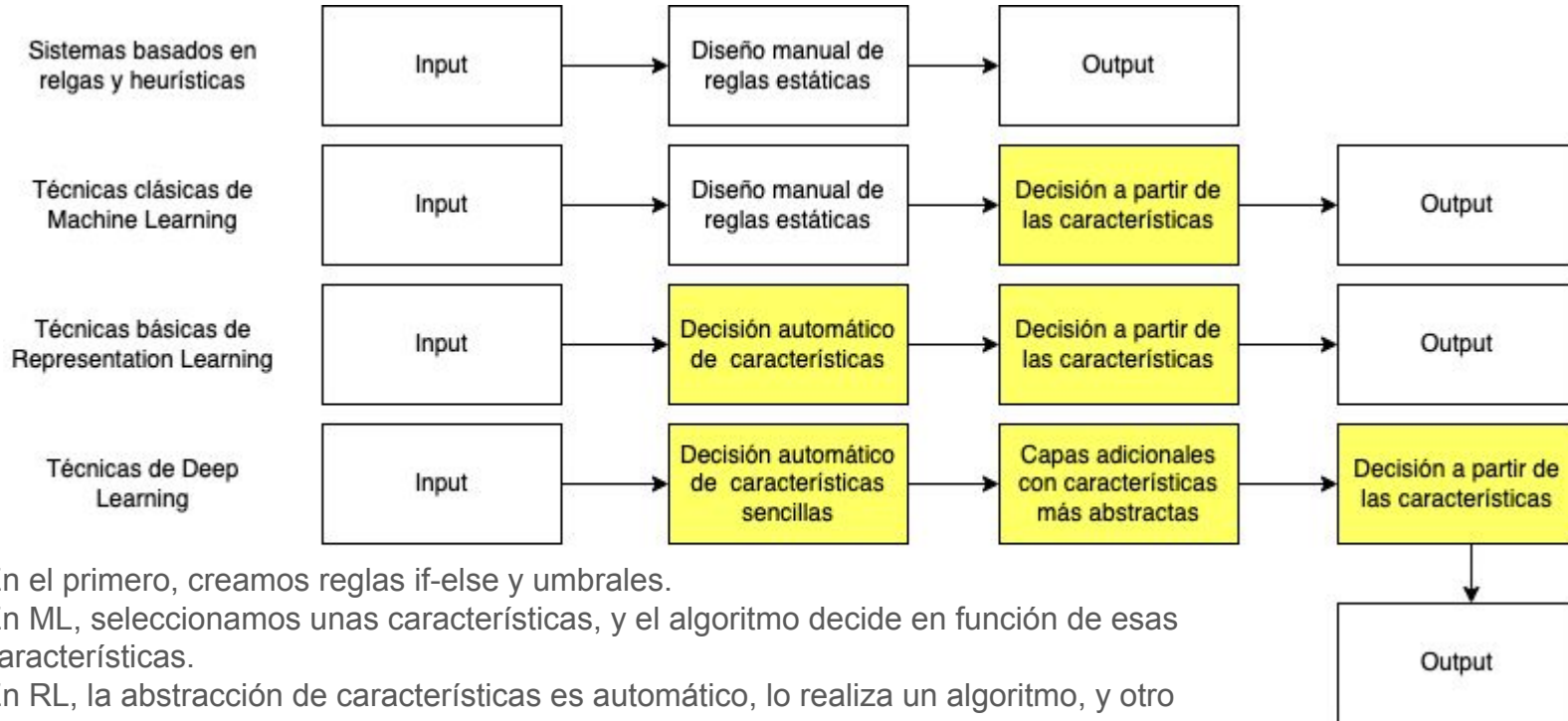


¿Qué es el Deep Learning?

Esto es solo una intuición, iremos viendo de manera teórica cómo funciona cada una de las neuronas artificiales y cada una de las capas.

Pero es importante tener esta idea preconcebida y, después, verlo formal y teóricamente de que cada una de las capas lo que van a hacer es ir ajustando una serie de parámetros, de manera que cada neurona artificial de esa capa se va a encargar de identificar una característica que aparece en la imagen. Y, normalmente, las primeras capas de la red serán características más básicas, más sencillas, que irán convergiendo y construyendo características más complejas a medida que avanzamos en las capas de nuestra red neuronal.

¿Qué es el Deep Learning? (en amarillo: sin analista)



En el primero, creamos reglas if-else y umbrales.

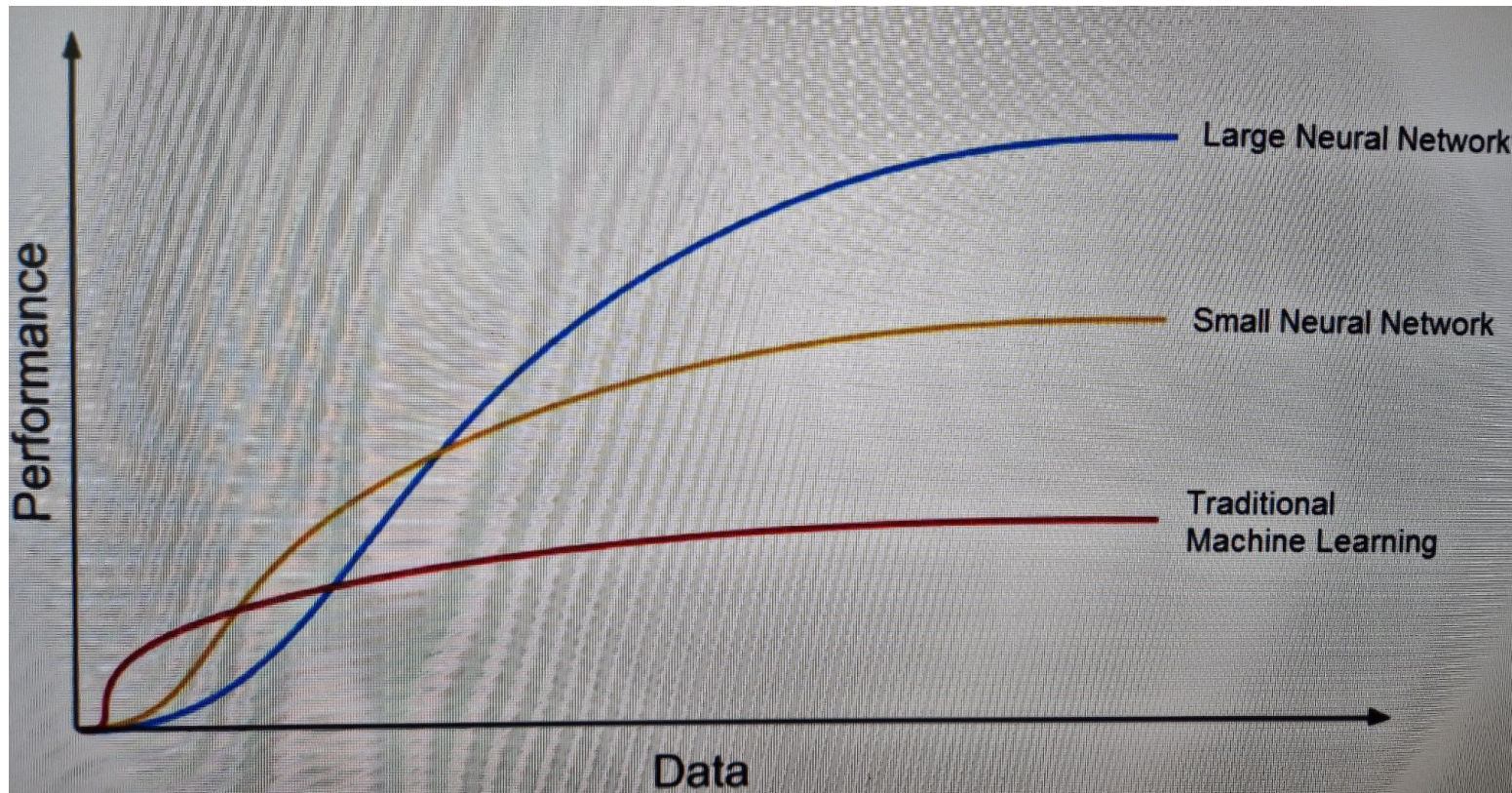
En ML, seleccionamos unas características, y el algoritmo decide en función de esas características.

En RL, la abstracción de características es automático, lo realiza un algoritmo, y otro algoritmo toma la decisión a partir de esas características.

En DL (E2E Learning), cuando la abstracción no funciona correctamente, empezará por un diseño automático de características sencillas que después convergerán en características más complejas, más abstractas. Para finalmente devolvernos el output.

Aplicaciones y clasificación de las técnicas de DL

Para que los algoritmos de DL funcionen adecuadamente, se requiere un conjunto de datos grande (para conjuntos pequeños, se comporta mejor ML):



Aplicaciones y clasificación de las técnicas de DL

Como hemos visto, todas aquellas técnicas clásicas de ML, siguen siendo de mucha utilidad en la actualidad. Porque muchas veces es complicado obtener un conjunto de datos que sea suficientemente grande (esto en deep learning, estamos hablando de un millón de ejemplos o más).

Si, por ejemplo, queremos identificar transacciones bancarias fraudulentas, necesitaríamos recopilar, al menos, un millón de transacciones pasadas etiquetadas como fraudulentas o legítimas.

Y, probablemente, si nosotros tenemos un conjunto de datos pequeño ($n < 100k$), una técnica de ML tradicional, como las que vimos en el 1Q, pueda servirnos y pueda darnos, incluso, mejores resultados que una red neuronal profunda.

Aplicaciones y clasificación de las técnicas de DL

Es importante siempre comprender dónde se encuentra aproximadamente estos puntos de corte (véase la gráfica anterior), porque seguramente en muchas ocasiones nos resulte más rentable (eficiente) utilizar técnicas que sean más sencillas, pero que para nuestro volumen de conjunto de datos, pues nos van a proporcionar mejores resultados y, además, más rápido.

- El Deep Learning ha conseguido resultados excelentes en áreas complejas que no podían resolverse adecuadamente con técnicas clásicas de ML
 - Anuncios dirigidos y marketing
 - Clasificación de imágenes
 - Reconocimiento de audio
 - Traducción automática avanzada
 - Conversión de texto a voz
 - Conducción autónoma
 - Simulación de fotos y vídeos, ...

Aplicaciones y clasificación de las técnicas de DL

Es importante siempre comprender dónde se encuentra aproximadamente estos puntos de corte (véase la gráfica anterior), porque seguramente en muchas ocasiones nos resulte más rentable (eficiente) utilizar técnicas que sean más sencillas, pero que para nuestro volumen de conjunto de datos, pues nos van a proporcionar mejores resultados y, además, más rápido.

- El Deep Learning ha conseguido resultados excelentes en áreas complejas que no podían resolverse adecuadamente con técnicas clásicas de ML
 - Anuncios dirigidos y marketing
 - Clasificación de imágenes
 - Reconocimiento de audio
 - Traducción automática avanzada
 - Conversión de texto a voz
 - Conducción autónoma
 - Simulación de fotos y vídeos, ...

Aplicaciones y clasificación de las técnicas de DL

¿Y qué tipo de redes neuronales profundas o qué tipo de algoritmos de Deep Learning van a solucionar, o van a comportarse adecuadamente, para este tipo de problemas?

- Anuncios dirigidos y marketing -> **Redes neuronales artificiales profundas**
- Clasificación de imágenes -> **Redes Neuronales Convolucionales (RNA específica)**
- Reconocimiento de audio
- Traducción automática avanzada -> **Redes Neuronales Recurrentes**
- Conversión de texto a voz
- Conducción autónoma
- Simulación de fotos y vídeos -> **Redes Generativas Antagónicas (JAMS, rel. nuevas)**

Nosotros nos centraremos en las primeras, las RNA profundas. Aunque el resto son también RNAs que tienen alguna particularidad específica que las hace comportarse mejor para estos problemas en concreto. Y la base de todas ellas es lo que vamos a tratar en este cuatrimestre.