



Universidad Tecnológica de Panamá
Facultad de Ingeniería Eléctrica
Maestría en Ingeniería Eléctrica

Semana 1

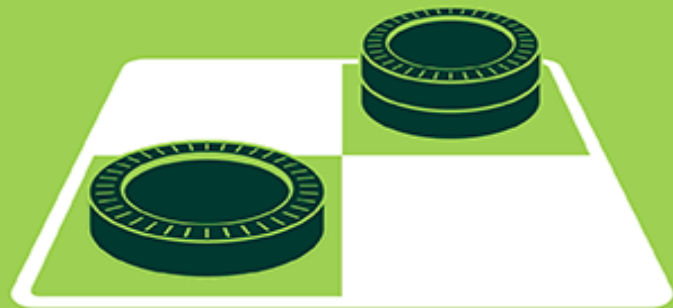
Introducción a Machine Learning

Definición

Machine Learning es una rama de la inteligencia artificial que permite a las computadoras aprender a partir de datos y mejorar su desempeño sin ser programadas explícitamente.

ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.



DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's

1960's

1970's

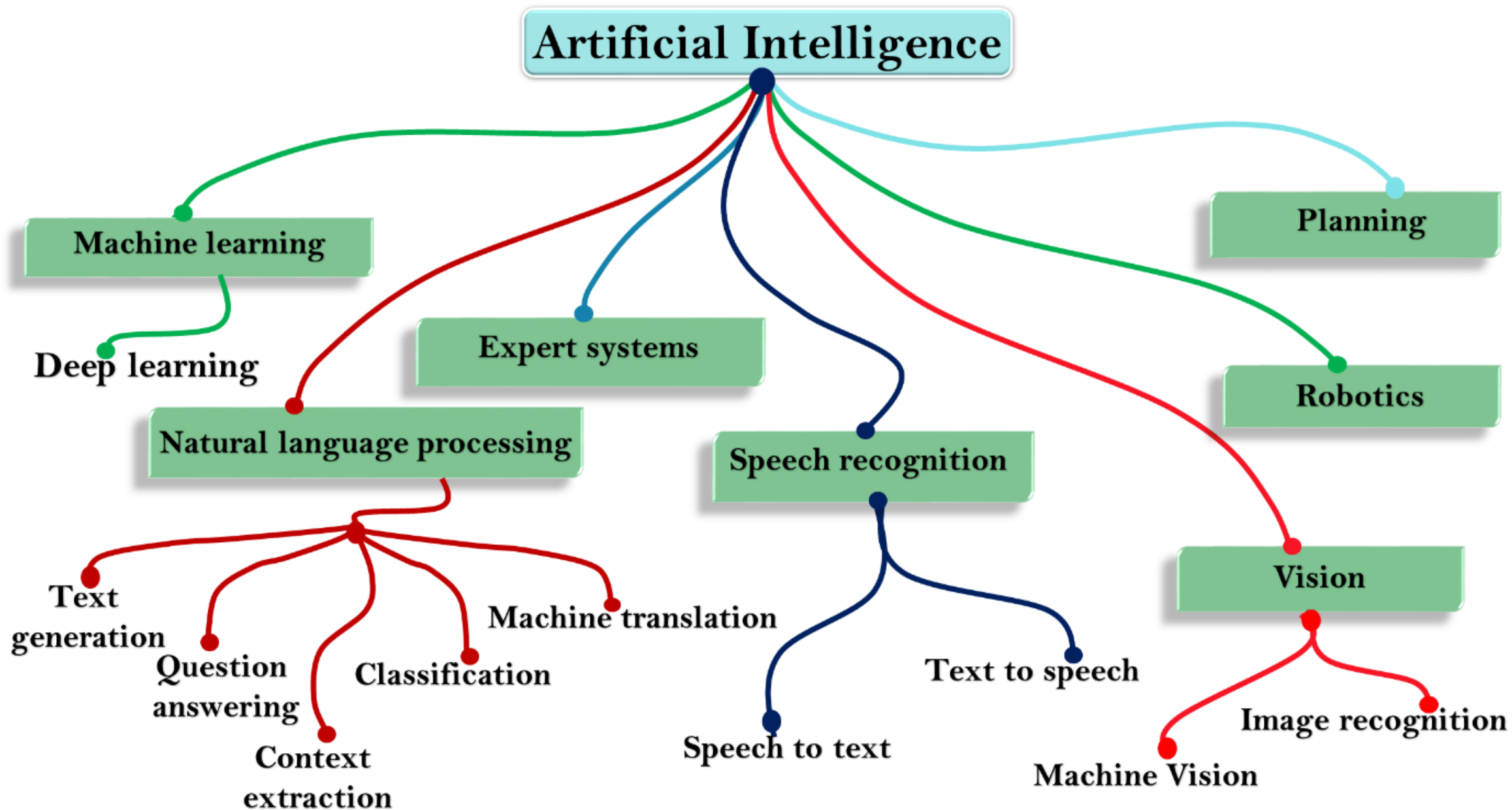
1980's

1990's

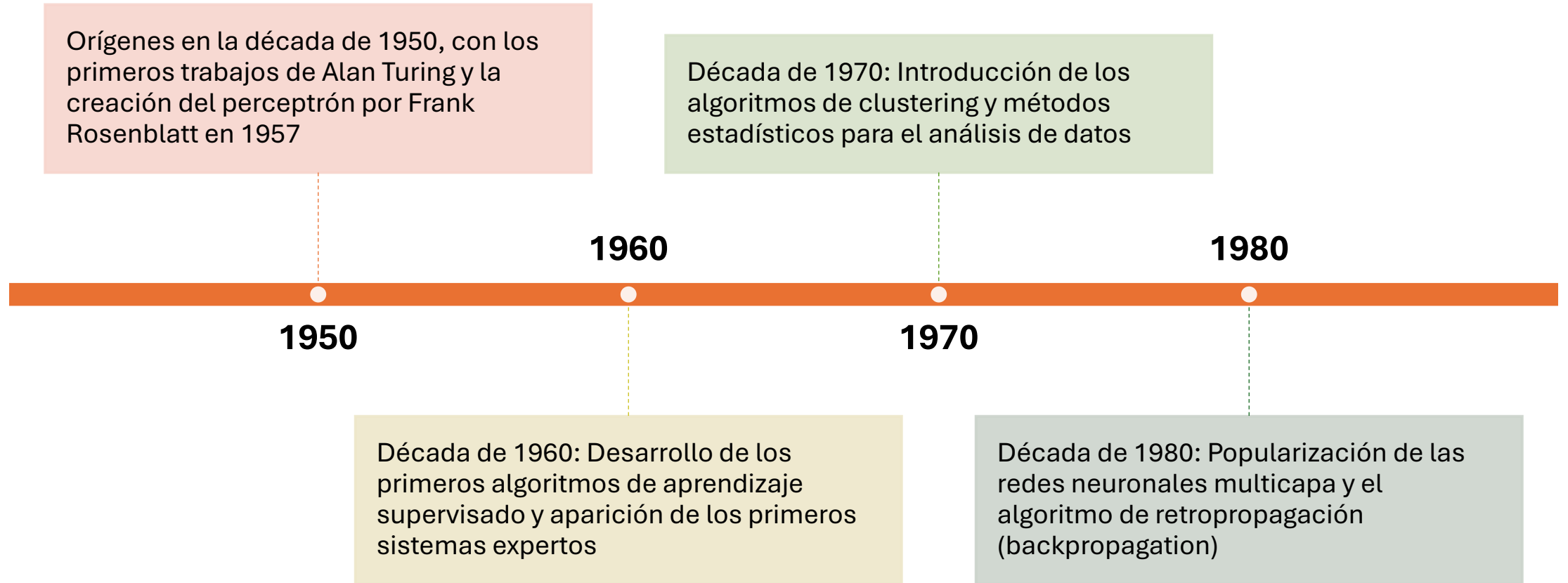
2000's

2010's

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



Historia y evolución



En la [historia de la inteligencia artificial](#), un **Invierno IA** es un período de reducción de fondos e interés en la investigación de [inteligencia artificial](#).¹ El término fue acuñado por [analogía](#) a la idea del [invierno nuclear](#).²

El término apareció por primera vez en 1984 como el tema central de un debate público en la conferencia anual de la [AAAI](#). Es una reacción en cadena que comienza con el pesimismo de la comunidad de IA, seguido por el pesimismo en la prensa, seguido de un severo recorte en la financiación, seguido por el final de la investigación seria.² En la conferencia, Roger Schank y [Marvin Minsky](#)—dos de los principales investigadores de la IA que habían sobrevivido el "invierno" de la década de 1970, advirtieron a la comunidad de negocios que el entusiasmo por la IA había crecido de forma descontrolada en la década de 1980 y que, sin duda, la decepción ciertamente seguiría. Tres años más tarde, la industria de la IA mil millones de dólares comenzó a derrumbarse.²

El furor es común en diversas tecnologías emergentes, como lo fue la [Mania del Ferrocarril](#) o la [Burbuja puntocom](#). El Invierno IA fue un resultado de ese furor, debido a promesas poco realistas por parte de los desarrolladores, expectativas altas de los usuarios finales y una amplia promoción en los medios.³ A pesar de la subida y la caída de la reputación de la IA, se ha continuado desarrollando nuevas tecnologías y exitosas tecnologías. El investigador [Rodney Brooks](#) se quejaría en 2002 de que "existe este estúpido mito de que la IA ha fallado, pero la IA esta a su alrededor cada segundo del día."⁴ En el 2005, [Ray Kurzweil](#) estaba de acuerdo: "Muchos observadores siguen pensando que el invierno IA fue el final de la historia y que nada desde entonces ha venido del campo IA. Sin embargo, hoy en día miles de aplicaciones de la IA están profundamente arraigados en la infraestructura de todas las industrias."⁵

Es entusiasmo y optimismo sobre la IA ha aumentado gradualmente desde su punto más bajo en 1990. A partir de la década del 2010 la [Inteligencia artificial](#) (y especialmente el subcampo del [Aprendizaje automático](#)) empezó a ganar interés por parte de la comunidad de investigación, lo que llevó a un auge dramático en el financiamiento y la inversión del sector.

Historia y evolución



Década de 1990: Avances en máquinas de soporte vectorial (SVM), árboles de decisión y métodos de ensemble como random forests.

Surgimiento del término "deep learning" y avances en redes neuronales profundas.

2000

1990

2006

Década de 2000: Expansión del aprendizaje automático en la industria, crecimiento de los datos digitales y aparición de grandes conjuntos de datos.

Historia y evolución



Aparición y adopción de la función de activación ReLU y el optimizador Adam, que permitieron entrenar redes profundas de manera más eficiente.

Desarrollo de frameworks como TensorFlow y PyTorch, y aplicaciones exitosas en visión por computadora, procesamiento de lenguaje natural y juegos.

2007–2010

2006–2012

2010

Uso creciente de GPUs para entrenamiento de modelos de deep learning y lanzamiento de CUDA por NVIDIA, acelerando el desarrollo de modelos complejos.

Historia y evolución



2017

Publicación de "Attention is All You Need", introduciendo el mecanismo de atención y los modelos Transformer, revolucionando el procesamiento de lenguaje natural y otras áreas.



2020-presente

Democratización de la inteligencia artificial a través de los modelos de lenguaje grande (LLM), permitiendo el acceso masivo a herramientas avanzadas de IA y facilitando su integración en productos y servicios cotidianos.



2022-presente

Avances recientes gracias a la disponibilidad de grandes volúmenes de datos (big data), mayor capacidad de cómputo (GPU/TPU), y algoritmos más sofisticados como el deep learning.

Provided proper attribution is provided, Google hereby grants permission to reproduce the tables and figures in this paper solely for use in journalistic or scholarly works.

Attention Is All You Need

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez*[†]
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaizer@google.com

Illia Polosukhin*[‡]
illia.polosukhin@gmail.com

Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

[†]Work performed while at Google Brain.

[‡]Work performed while at Google Research.

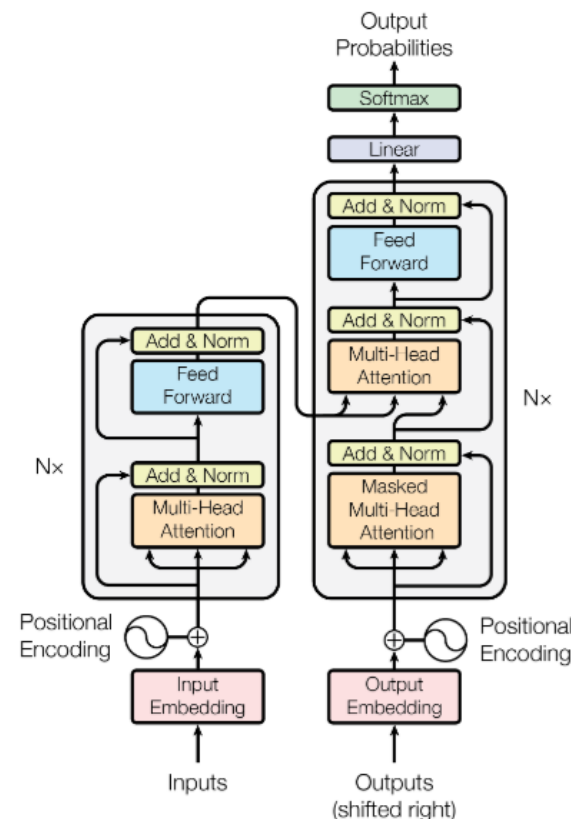


Figure 1: The Transformer - model architecture.

Attention is all you need

[A Vaswani](#), [N Shazeer](#), [N Parmar](#)... - Advances in neural ..., 2017 - proceedings.neurips.cc

... to attend to **all** positions in the decoder up to and including that position. **We need** to prevent ... **We** implement this inside of scaled dot-product **attention** by masking out (setting to $-\infty$) ...

☆ Guardar 📄 Citar Citado por 179467 Artículos relacionados Las 73 versiones 🔗

Impacto actual

Machine Learning está transformando industrias enteras, permitiendo automatizar tareas complejas, mejorar la toma de decisiones y crear productos y servicios innovadores.



Impacto actual

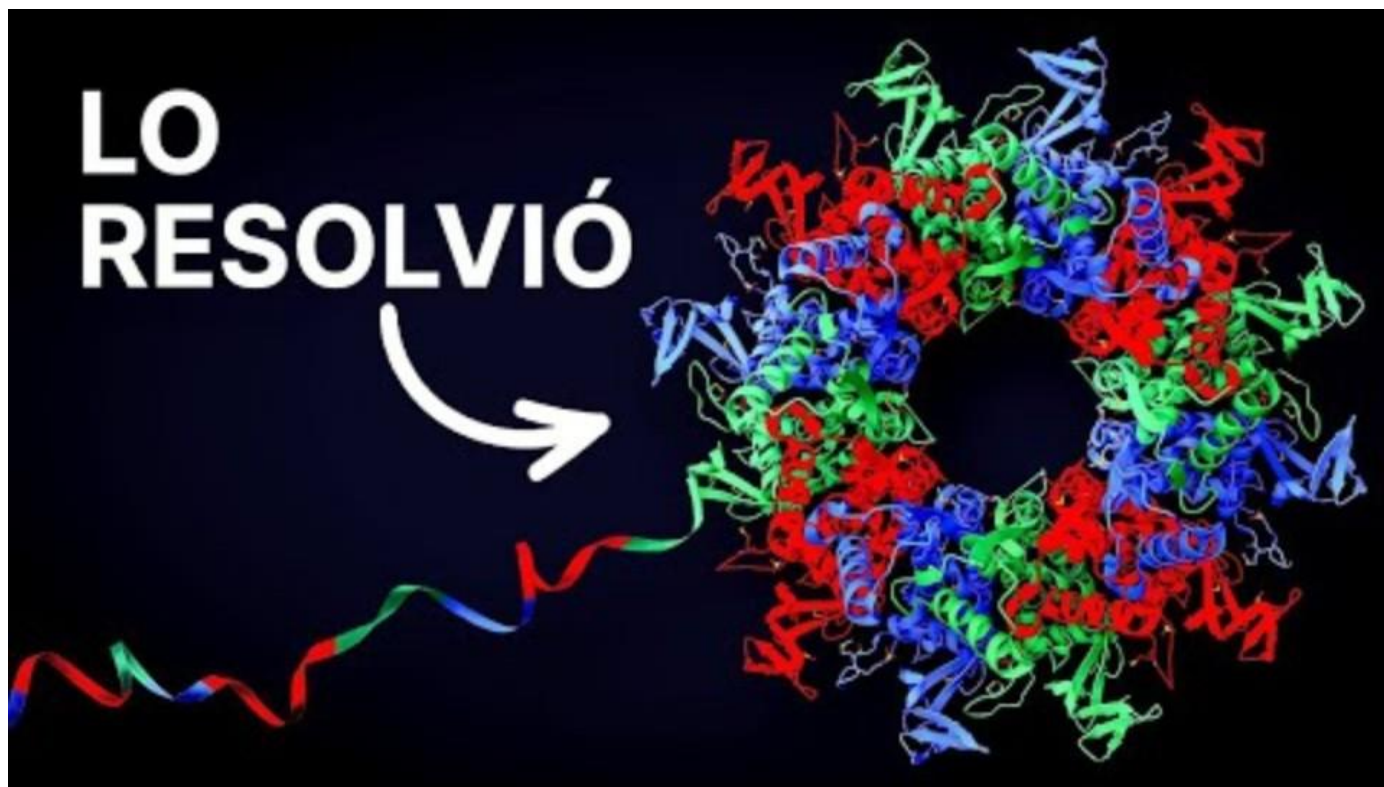
- En la industria, ha permitido optimizar procesos de manufactura, reducir costos y mejorar la calidad de productos mediante el monitoreo inteligente y el mantenimiento predictivo.
- En el sector salud, ha revolucionado el diagnóstico médico, la personalización de tratamientos y la investigación de nuevos medicamentos.
- En el ámbito financiero, ha mejorado la detección de fraudes, la gestión de riesgos y la toma de decisiones de inversión.



Impacto actual

- En la vida cotidiana, se encuentra presente en asistentes virtuales, sistemas de recomendación, traducción automática y reconocimiento de voz e imagen.
- En la ingeniería, ha impulsado el desarrollo de sistemas autónomos, la gestión eficiente de recursos energéticos y la simulación avanzada de sistemas físicos.
- Además, la aparición de los Large Language Models y herramientas accesibles ha democratizado el acceso a la inteligencia artificial, permitiendo que empresas, instituciones educativas y usuarios individuales puedan aprovechar sus beneficios en una amplia variedad de aplicaciones.





**Video
recomendado**

Aplicaciones

- Optimización del consumo energético
- Diagnóstico médico asistido por computadora
- Predicción de fallas en sistemas mecánicos
- Vehículos autónomos y sistemas de navegación inteligente
- Control inteligente de procesos de manufactura
- Recomendaciones de productos en plataformas de comercio electrónico y streaming
- Análisis de calidad de energía

Aplicaciones

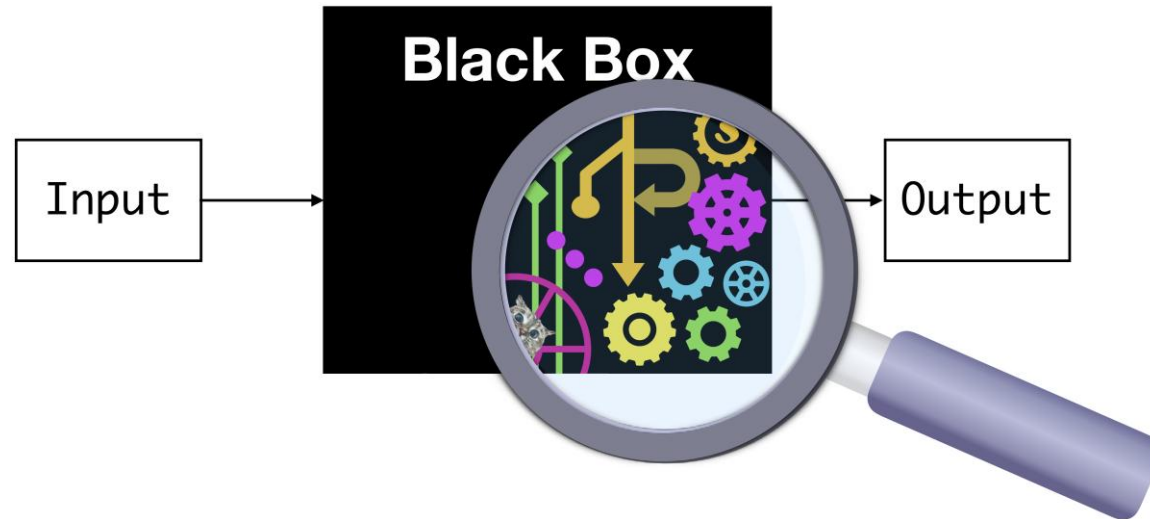
- Reconocimiento de voz e imagen (por ejemplo, asistentes virtuales y sistemas de seguridad)
- Mantenimiento predictivo en sistemas eléctricos
- Monitoreo y diagnóstico de maquinaria
- Optimización de diseño de componentes mecánicos
- Detección de fallas en redes eléctricas
- Detección de fraudes en transacciones financieras
- Detección temprana de anomalías en sistemas industriales complejos
- Pronóstico de demanda eléctrica y gestión de redes inteligentes (smart grids)

Aplicaciones

- Supervisión automática de procesos de fabricación para mejorar la calidad y reducir desperdicios
- Diagnóstico predictivo de motores eléctricos y sistemas de transmisión mecánica
- Optimización de rutas y logística en sistemas de transporte automatizados
- Análisis de vibraciones y señales para mantenimiento preventivo en maquinaria pesada
- Modelado y simulación de sistemas físicos complejos para diseño y control avanzado

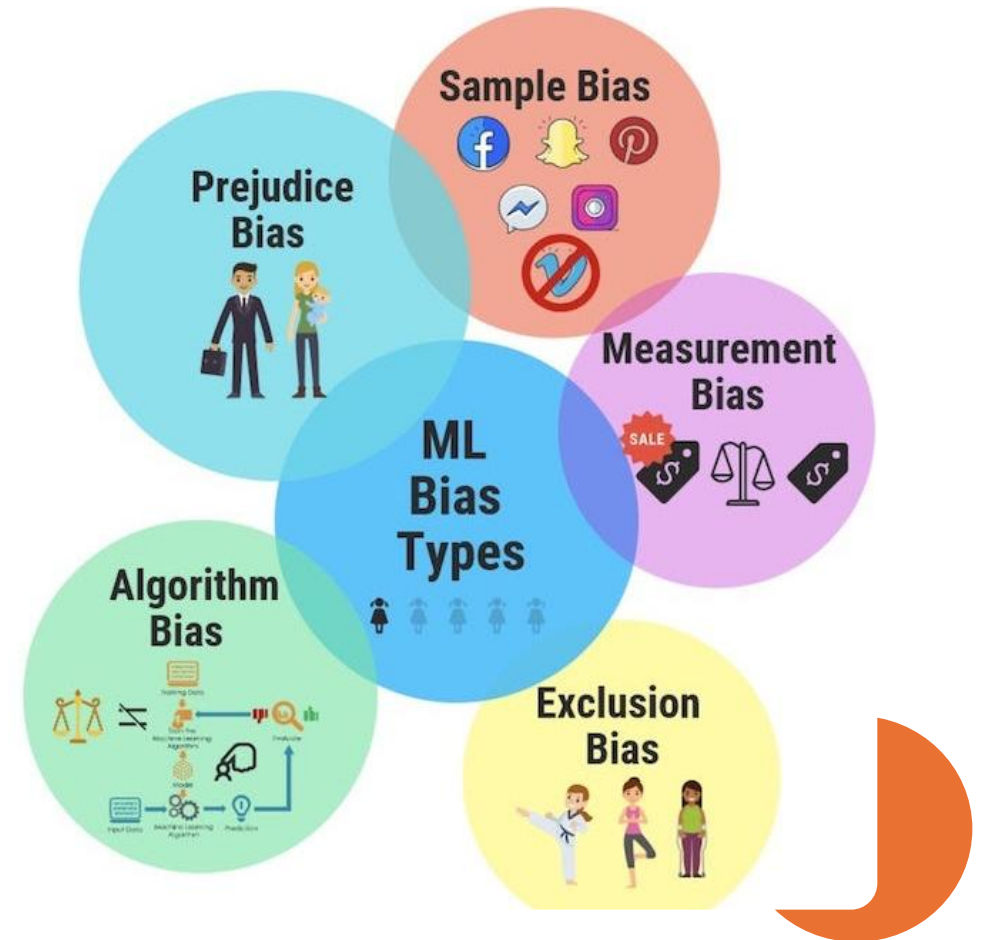
Desafíos actuales

Interpretabilidad y transparencia de los modelos: comprender cómo y por qué un modelo toma ciertas decisiones sigue siendo un reto, especialmente en modelos complejos como las redes neuronales profundas.



Desafíos actuales

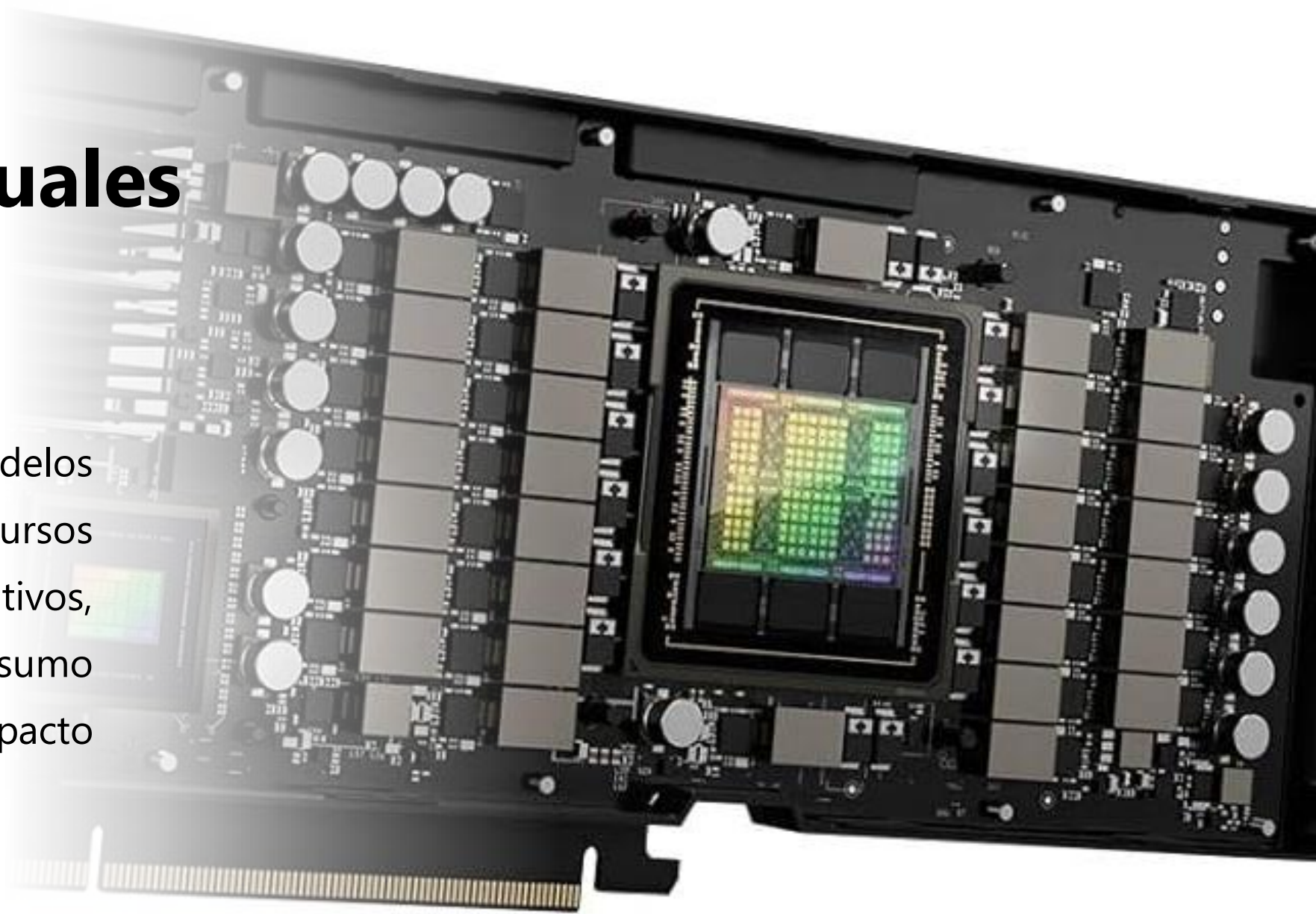
Sesgo y equidad en los datos y algoritmos: los modelos pueden perpetuar o amplificar sesgos presentes en los datos de entrenamiento, afectando la equidad y la justicia en las decisiones automatizadas.



Desafíos actuales

Sostenibilidad y consumo energético

El entrenamiento de modelos grandes requiere recursos computacionales significativos, lo que implica un alto consumo energético y un impacto ambiental.



How much is the NVIDIA H100 in dollars?

The NVIDIA H100's price in dollars varies:

- **Base Price:** ~\$25,000
- **Advanced Configurations (e.g., H100 80GB):** ~\$30,970.79 The price depends on the model, memory size, and vendor-specific markup.

How many NVIDIA H100 GPUs does Tesla use?

Tesla is reported to have deployed around **35,000 NVIDIA H100 GPUs** in its private cloud. This makes Tesla one of the largest adopters of NVIDIA H100 GPUs, primarily for AI and self-driving research.

Can I lease an NVIDIA H100 GPU?

Yes, leasing is a popular option for enterprises. Many providers, such as **Jarvislabs**, **Lambda Labs**, and **RunPod**, offer NVIDIA H100 GPUs on a pay-as-you-go basis. This approach eliminates upfront hardware costs while allowing flexible scaling.

What is the power consumption of the NVIDIA H100 GPU?

The NVIDIA H100 GPU requires up to **700W of power under full load**. For multi-GPU setups, additional power distribution and cooling infrastructure are necessary to support the high energy requirements.

Recursos de hardware recomendados para ejecutar DeepSeek de manera local

Modelo	Tamaño	VRAM Requerida (4-bit)	GPU Recomendada
DeepSeek-R1-Distill-Qwen-1.5B	1.5B	~1 GB	NVIDIA RTX 3050 8GB o superior
DeepSeek-R1-Distill-Qwen-7B	7B	~4 GB	NVIDIA RTX 3060 12GB o superior
DeepSeek-R1-Distill-Llama-8B	8B	~4.5 GB	NVIDIA RTX 3060 12GB o superior
DeepSeek-R1-Distill-Qwen-14B	14B	~8 GB	NVIDIA RTX 4080 16GB o superior
DeepSeek-R1-Distill-Qwen-32B	32B	~18 GB	NVIDIA RTX 4090 24GB o superior
DeepSeek-R1-Distill-Llama-70B	70B	~40 GB	Configuración multi-GPU (p. ej., NVIDIA RTX 4090 24GB x2)
DeepSeek-R1-Zero	671B	~1,342 GB	Configuración multi-GPU (p. ej., NVIDIA A100 80GB x16)
DeepSeek-R1	671B	~1,342 GB	Configuración multi-GPU (p. ej., NVIDIA A100 80GB x16)

Probando distintos modelos de DeepSeek

```
PS C:\Windows\System32> ollama run deepseek-r1:1.5b
```

```
>>> explícame, qué es Panamá?
```

```
<think>
```

```
</think>
```

Panamaría (o se pronuncia Panámaría) es una tienda de coserías y madera que se encuentra en el corazón de la ciudad de Panamá. Fue Established por dos foresters, el del Artesano y el del Aguañala, y ha ganado un populares throughout su history.

Historia

- ****Inicio****: Nace en 1927 como una zona rural de las ciudades de Panama.
- ****Ganación de importancia****: En 1964, fue convertida en una zona rural importante por el gobierno chino y severalias demencias de los States Unidos.
- ****Pared rectangular****: Posteriormente, el tesoro de esta zona, conocido como la "pared rectangular", incluye tiendas de coserías como Panamá, Vistamaría, Vinculos y Céntaro.

Funcionalidades

- ****Costumer****: Ideal para personas que quieren coser su cosita, madera o otros articles de moda.
- ****Sistema de financiamiento****: En un nivel microfinanciaro, pero puede ser ampliado con la adopción de applets para las personas más respetadas.

Panamá es una tienda destacada de la ciudad de Panama, con su gran importancia y amplio servicio al populares.

```
>>> Send a message (/? for help)
```

Probando distintos modelos de DeepSeek

```
PS C:\Windows\System32> ollama run deepseek-r1:1.5b
```

```
>>> what is Panama?
```

```
<think>
```

```
</think>
```

Panama is a country in South America, located in the southern part of the country. It shares borders with Colombia to the west and Venezuela to the east. The capital city is Panamericana.

Panama was an independent country before 1959 when it joined the United Nations as a member of its regional organization, the Dominican Republic (DR), Ecuador, and Mexico. After its membership in 1959, Panama became a permanent United Nations member state. The government of Panama is led by President Juan Manuel Díaz de León.

Panama has been one of the most prosperous countries in Latin America, with a strong agricultural economy, especially in the regions of Panama City (which includes Panama City, El Central and another city). Panamanian culture has also had a significant impact on neighboring South American countries.

```
>>> Send a message (/? for help)
```


Probando distintos modelos de DeepSeek



Probando distintos modelos de DeepSeek

≡ deepseek-r1:70b

Economía

- Panamá tiene una economía relativamente estable en comparación con otros países de América Central.
- Su moneda oficial es el balboa, pero el dólar estadounidense también se usa ampliamente.
- El Canal de Panamá es una fuente importante de ingresos para el país, junto con el turismo, la agricultura, la pesca y los servicios financieros.

Cultura

- La cultura panameña es un mestizaje de influencias indígenas, españolas, afroantillanas y estadounidenses.
- El país celebra numerosas fiestas tradicionales, como el Carnaval en Panamá, que es uno de los más famosos de América Latina.
- La gastronomía incluye platos como el arroz con pollo, la sancocho (un estofado), las empanadas y los tamales.

Biodiversidad

- A pesar de su tamaño relativamente pequeño, Panamá alberga una gran diversidad de flora y fauna.
- Es hogar de parques nacionales como Soberanía, Chagres y Barú, donde se pueden encontrar jaguares, monos, aves exóticas y más.

Política

- Panamá es una república democrática con un sistema presidencialista. El presidente es el jefe de Estado y gobierno.
- Desde la caída del régimen militar en 1989, el país ha mantenido una estabilidad política relativamente alta.

Panamá es un lugar único que combina modernidad, historia y naturaleza, lo que lo hace muy atractivo tanto para turistas

+ Send a Message

?



Desafíos actuales


Regulación y ética: la rápida adopción de la IA plantea desafíos regulatorios y éticos, como la responsabilidad ante errores, la transparencia en el uso de algoritmos y el respeto a los derechos humanos.

Privacidad y seguridad de la información: proteger los datos personales y sensibles utilizados para entrenar modelos, así como evitar ataques adversariales que puedan manipular los resultados.



Desafíos actuales

Actualización y mantenimiento de modelos: los modelos pueden volverse obsoletos si los datos cambian con el tiempo, por lo que es necesario actualizarlos y monitorearlos continuamente.



Accesibilidad y democratización: garantizar que los avances en IA estén disponibles para una amplia variedad de usuarios y sectores, evitando la concentración de poder tecnológico en pocas manos.

Entrando en materia

Clasificación vs Regresión

Regresión

Predice valores continuos, es decir, variables numéricas que pueden tomar cualquier valor dentro de un rango.

- Ejemplo: Predecir el precio de una casa, la temperatura de una ciudad, el ingreso anual de una persona.

Clasificación

Asigna categorías o etiquetas a los datos, es decir, predice clases discretas.

- Ejemplo: Clasificar correos como spam o no spam, diagnosticar si un paciente tiene una enfermedad (sí/no), identificar el tipo de flor según sus características.

Diferencias clave

- Regresión: salida numérica continua.
- Clasificación: salida categórica/discreta.



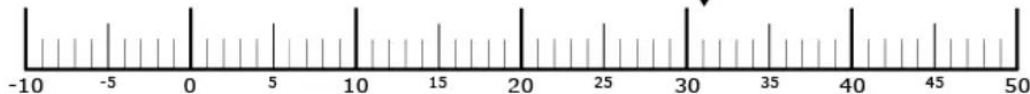
Regression

What temperature will there be tomorrow?

PREDICTION

31°

Celsius
°C



Classification

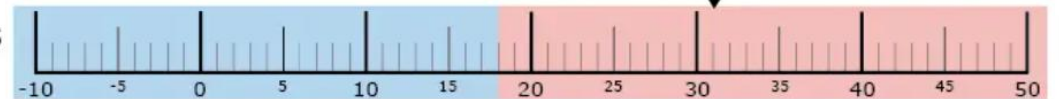
Will tomorrow be a cold or hot day?

Representación gráfica de la dif...

PREDICTION

Hot

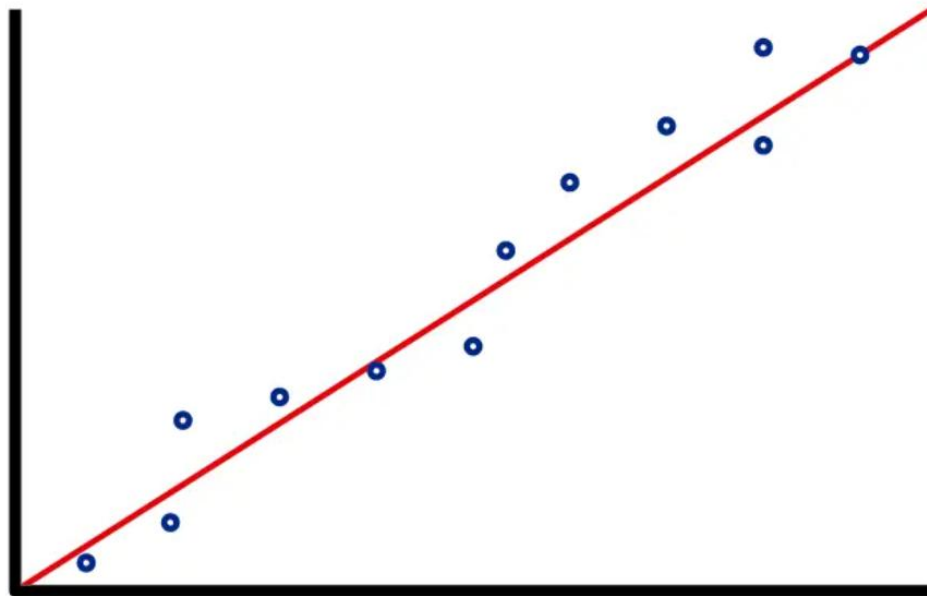
Celsius
°C



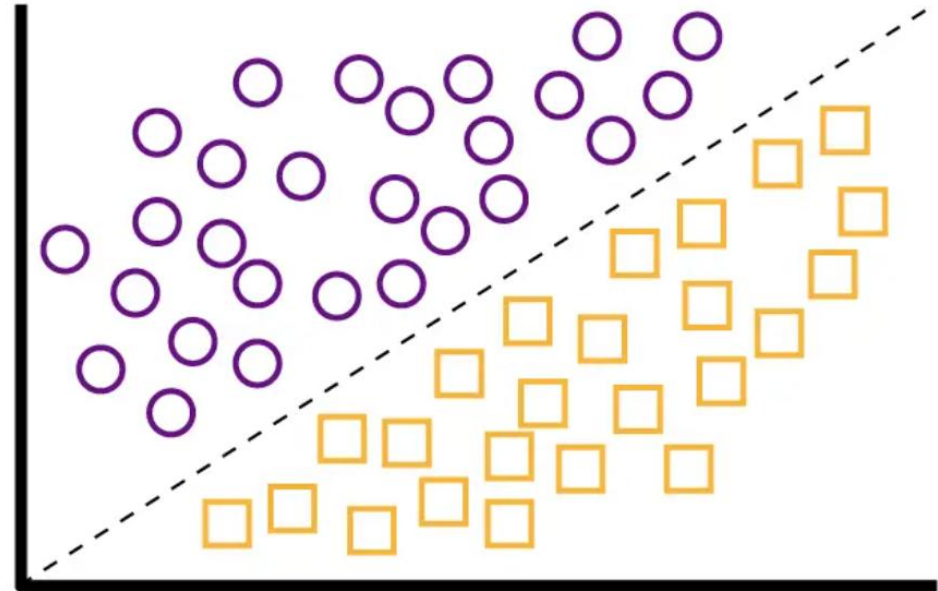
Diferencias clave

En regresión, los datos suelen representarse en gráficos de dispersión donde la variable objetivo es continua.

En clasificación, los datos pueden visualizarse en gráficos donde los puntos se agrupan por clases o colores.



Regression



Classification

Casos de uso

- **Regresión:** predicción de ventas, temperatura, demanda energética, precios de acciones.
- **Clasificación:** diagnóstico de enfermedades, reconocimiento de imágenes, detección de fraudes, análisis de sentimientos.