
Ciencia de Datos y Modelos Analíticos

A large, faint, light blue watermark of the BDA logo is centered in the background of the slide. It consists of a large 'BDA' acronym and the words 'BIG DATA ACADEMY' in a smaller font below it.

Ciencia de Datos

A large, faint, light blue watermark is centered on the slide. It consists of a stylized cloud icon with a mountain peak inside, followed by the text "BDA" in a large, bold, sans-serif font, and "BIG DATA ACADEMY" in a smaller, sans-serif font below it.

Ciencia de Datos

Extracción de conocimiento o de **patrones no triviales**, previamente desconocidos y **potencialmente útil**, sobre grandes bases de datos.



Objetivo

Búsqueda de patrones por medio
de :

Análisis descriptivo

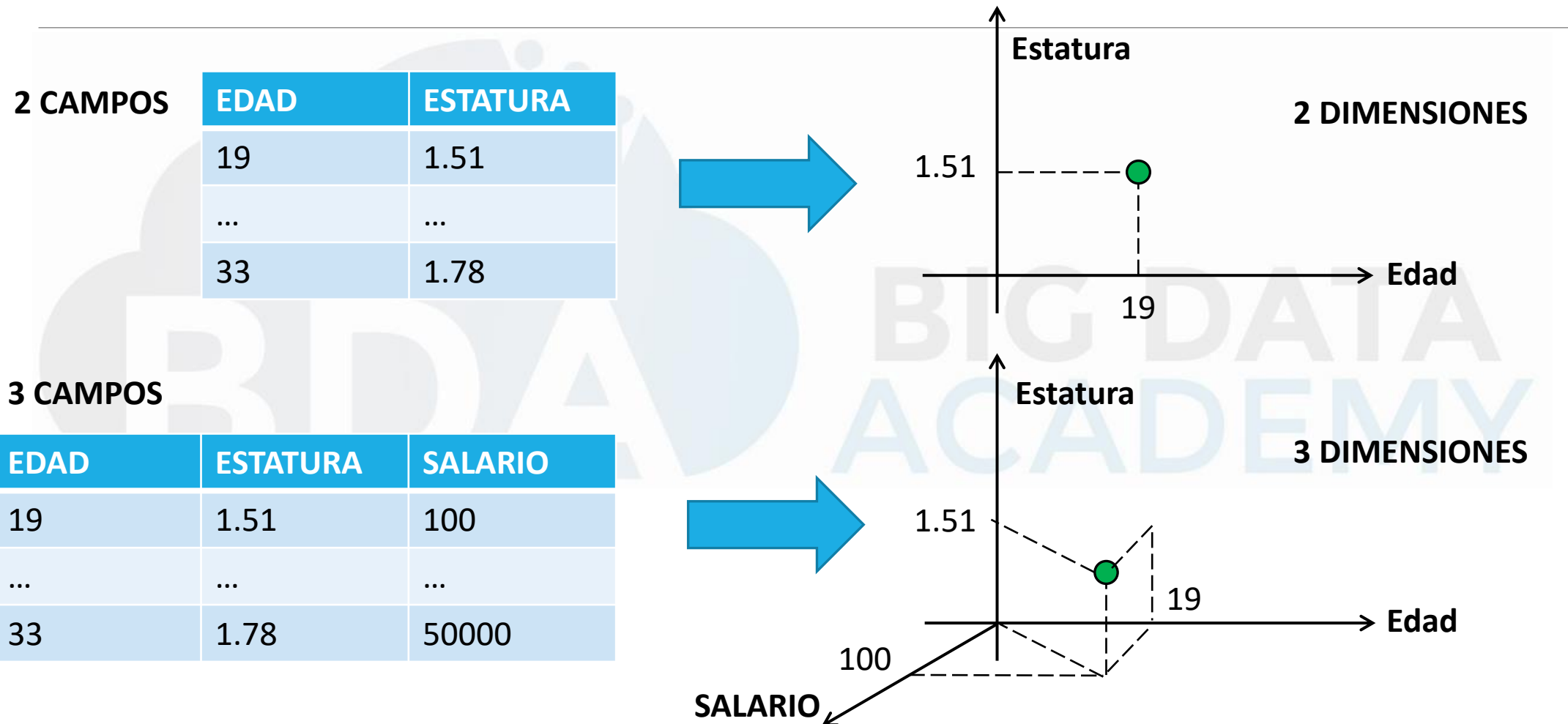
Análisis predictivo

Modelos multi- dimensionales

A large, faint, light blue watermark is centered in the background of the slide. It consists of a circular emblem on the left containing a stylized mountain and a network of nodes, and the text "BIG DATA ACADEMY" in a large, bold, sans-serif font to its right.

Puntos en N-Dimensiones

Un registro de una tabla de datos puede ser visto como un punto en un plano de N-Dimensiones



Tablas de más de 3 dimensiones

Tabla de N-Campos

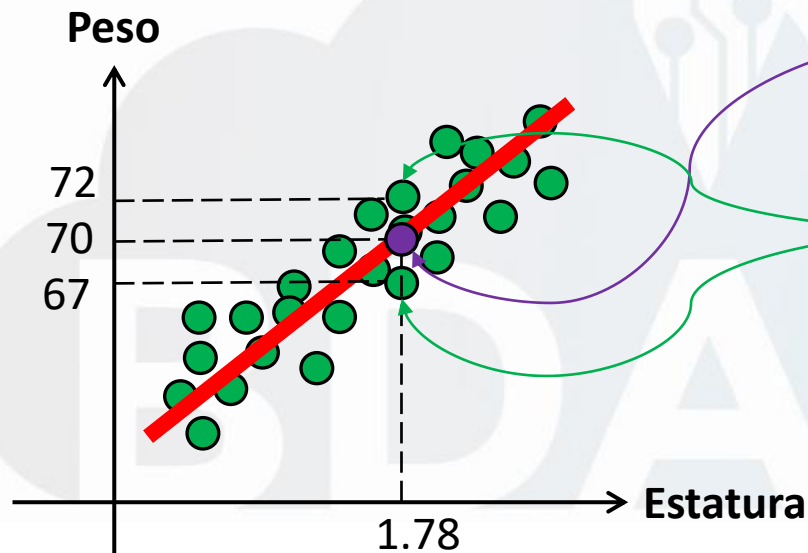
EDAD	ESTATURA	SALARIO	...	PESO
19	1.51	100	...	45
...
33	1.78	50000	...	70



Punto en N-Dimensiones

No puede representarse gráficamente, se analiza por medio de álgebra lineal y ecuaciones

Un modelo analítico es una idealización de la realidad



Modelo representando una realidad (**nube de puntos**)

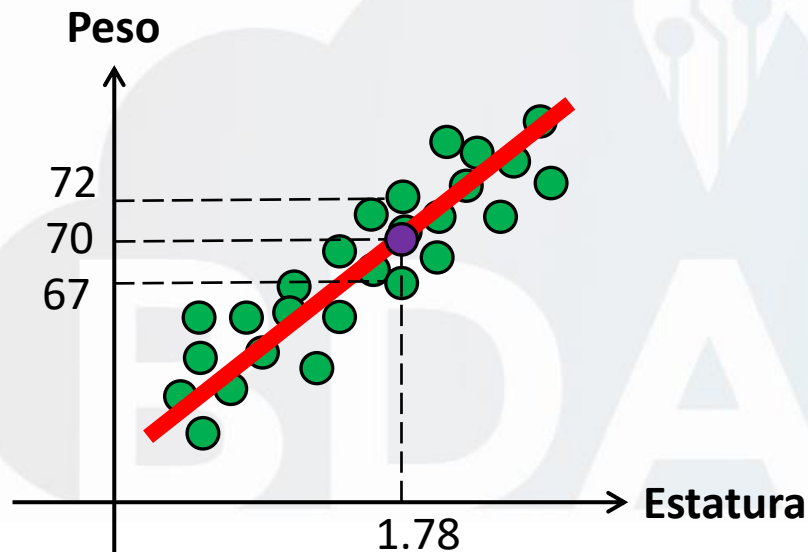
El modelo dice “Todas las personas de 1.78 Mt. pesan 70 Kg.”

En la realidad hay personas de 1.78Mt. Que pesan 72 Kg, 68 Kg, etc.

Realmente el modelo dice: “En promedio, las personas de 1.78 Mt. aproximadamente pesan 70 Kg.”

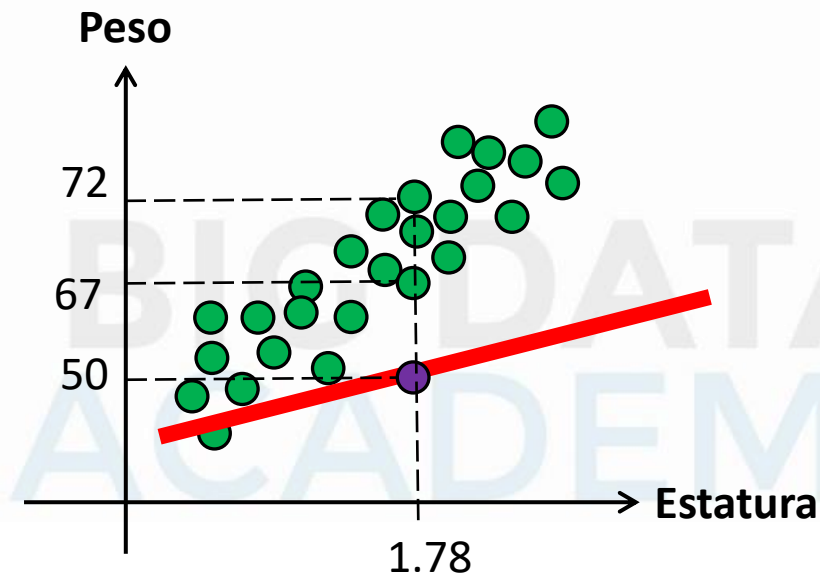
Objetivo del modelo: Minimizar el error lo más que se pueda

Minimización del error



Buen modelo: Se parece a la realidad

“En promedio, las personas de 1.78 Mt. aproximadamente pesan 70 Kg.”



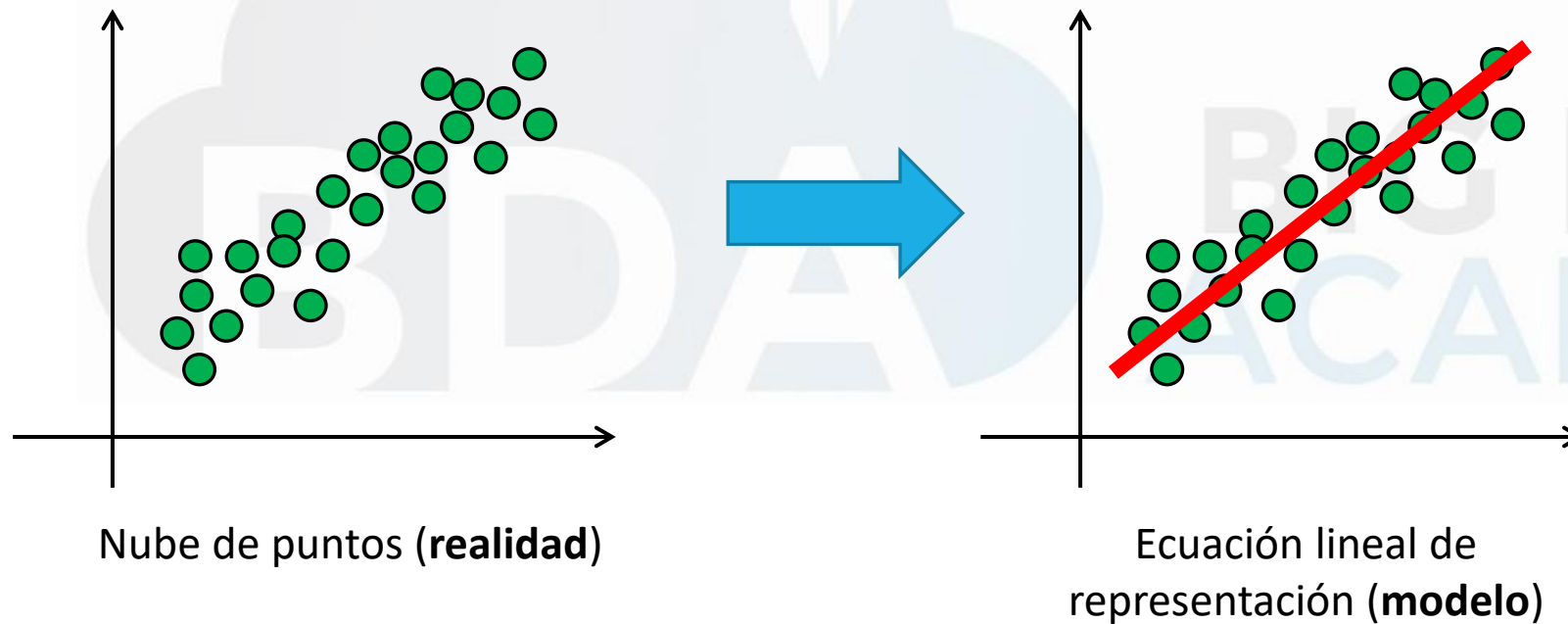
Mal modelo: No se parece a la realidad

“En promedio, las personas de 1.78 Mt. aproximadamente pesan 50 Kg.”

Tipos de Algoritmos Analíticos para crear modelos

Algoritmos de Machine Learning

Modela una realidad (nube de puntos) por medio de ecuaciones lineales



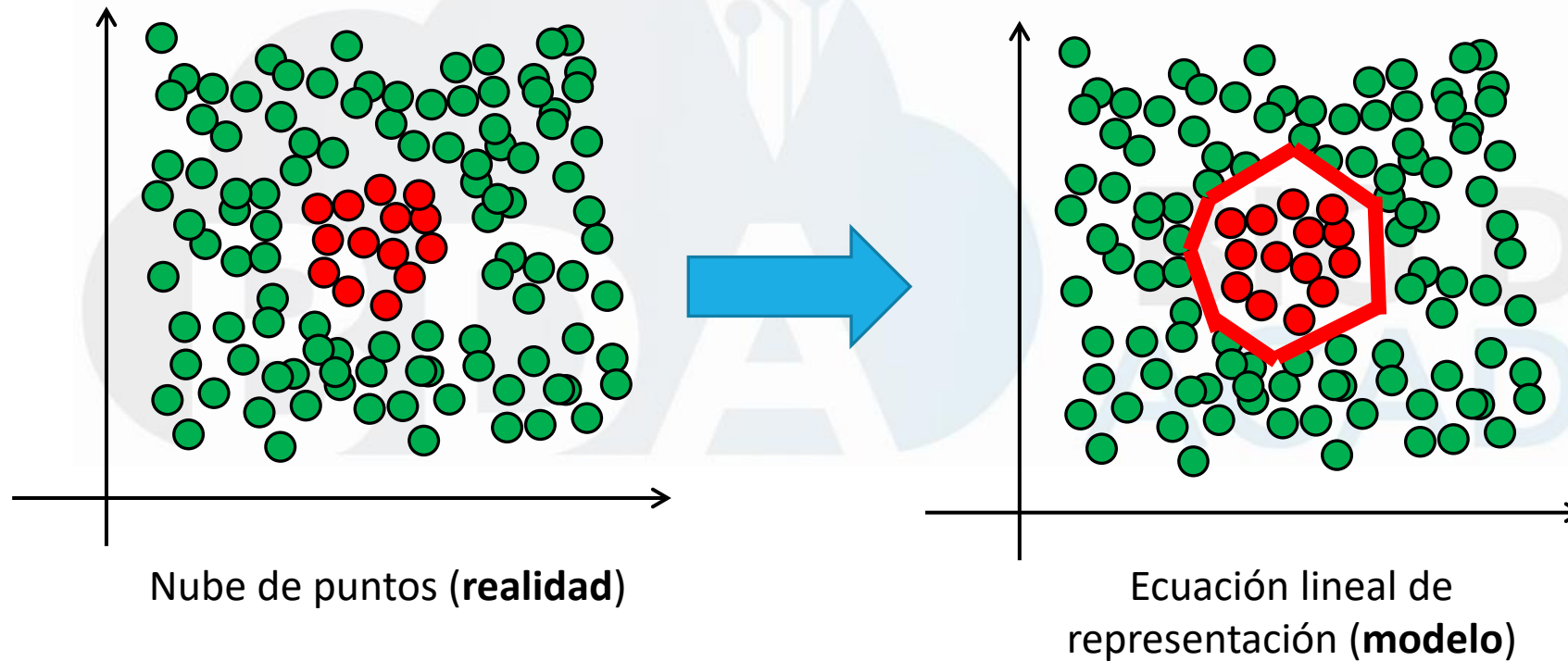
Dimensiones	Ecuación
2	Recta
3	Plano
4 o más	Hiper-cubo

Tipos de ecuaciones
lineal en N-Dimensiones

Se especializa en el descubrimiento de patrones lineales

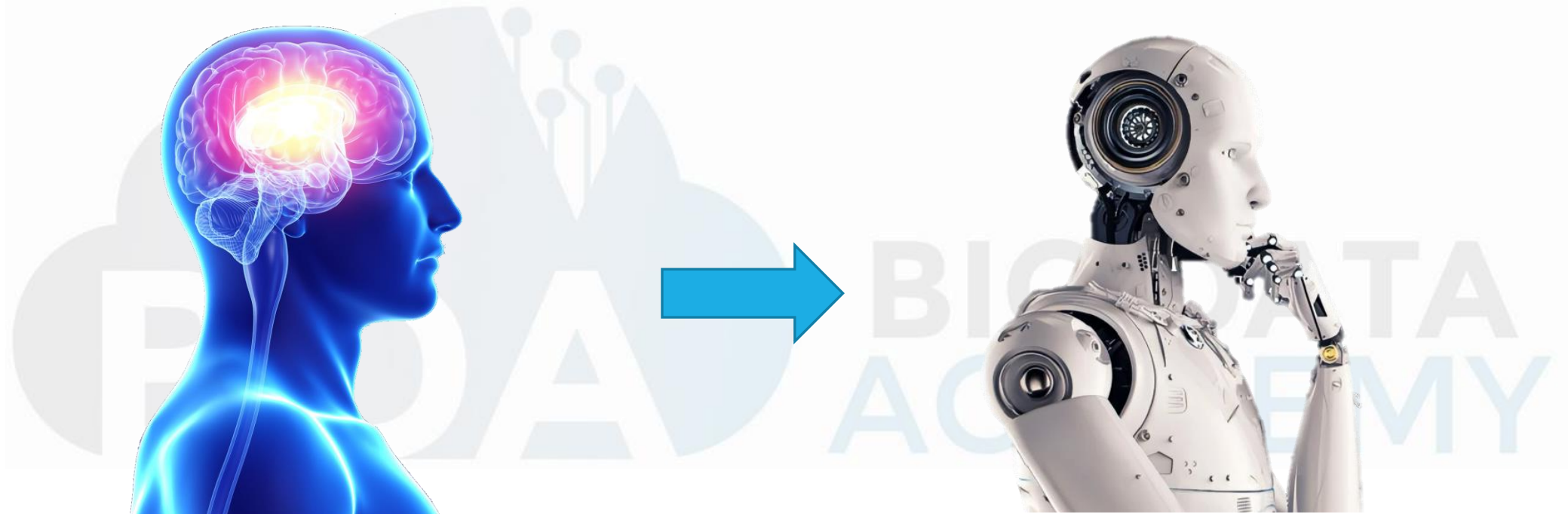
Algoritmos de Deep Learning

Modela una realidad no lineal (nube de puntos) por medio de varias ecuaciones lineales



Se especializa en el descubrimiento de patrones NO lineales

Algoritmos de Inteligencia Artificial



Inteligencia humana = Conocimiento + Experiencia a lo largo de muchos años

Inteligencia artificial = Conocimiento + Experiencia adquirida a mucha velocidad (días o meses)

Se especializa en emular el proceso de pensamiento humano para la toma de decisiones