

# APRENDIZAJE ESTADÍSTICO UNA INTRODUCCIÓN

Mario I. Caicedo

14 de mayo de 2022



HYPE

SPOTLIGHT ON BIG DATA

# Spotlight

**ARTWORK** Tamar Cohen, *Andrew J Buboltz*  
2011, silk screen on a page from a high school  
yearbook, 8.5" x 12"

# Data Scientist: *The Sexiest Job of the 21st Century*

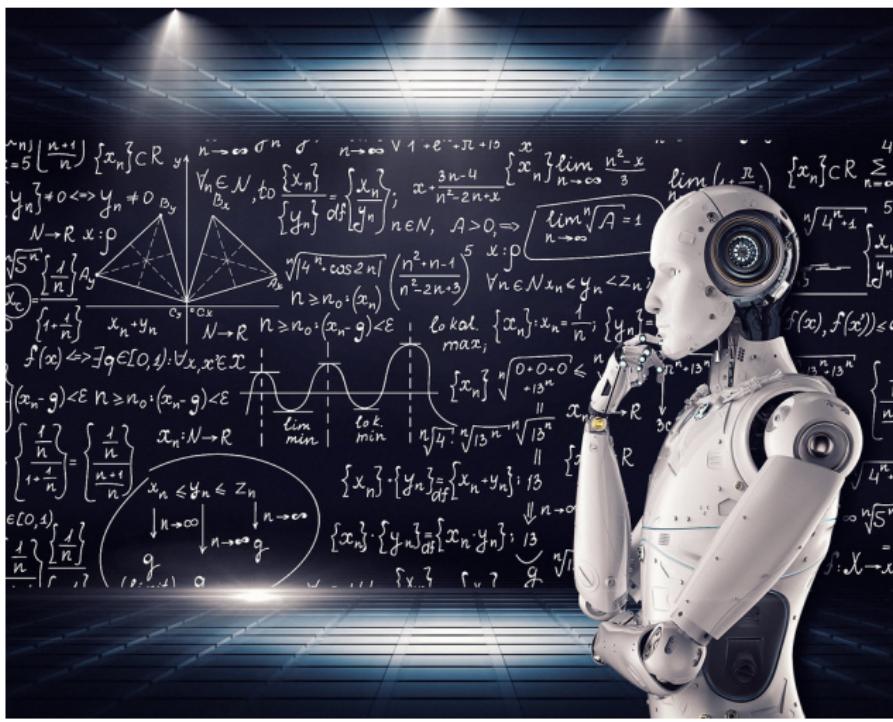
## **Meet the people who can coax treasure out of messy, unstructured data.**

70 Harvard Business Review October 2012





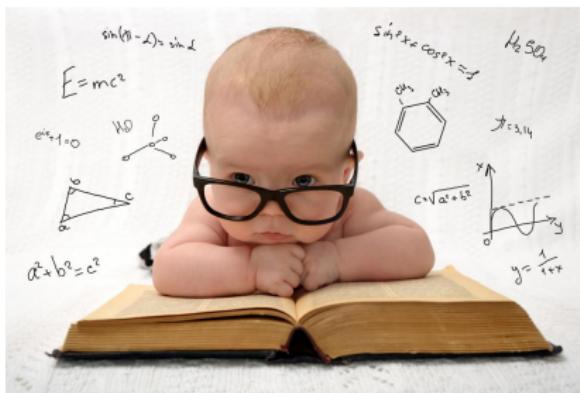
INTELIGENIA ARTIFICIAL (IA, AI)





# ¿COMO APRENDEMOS?

- Observamos.
- Sistematizamos las observaciones
- Notamos Patrones.
- Preguntamos acerca de tales patrones.
- Respondemos (de ser posible)
- Creamos un modelo.  
Síntesis de Información.
- Nuevas observaciones basadas en la síntesis y reinicio del proceso.

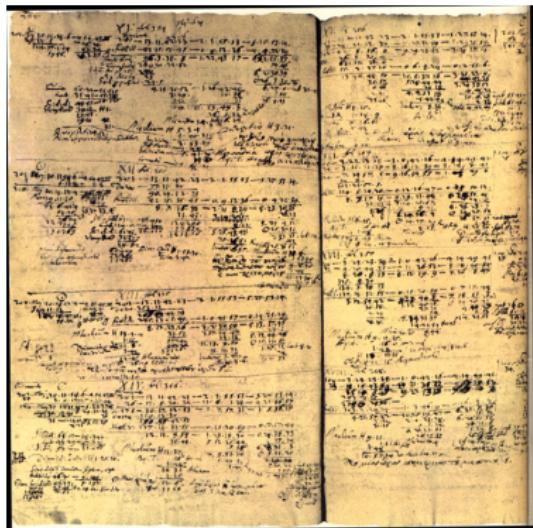


# ¿CALENDARIOS SOLARES?

Miles de años de observación del cielo (toma de datos) llevaron a la maravillosa síntesis de información alcanzada por Stonehenge. En la ilustración, el amanecer durante el solsticio de verano



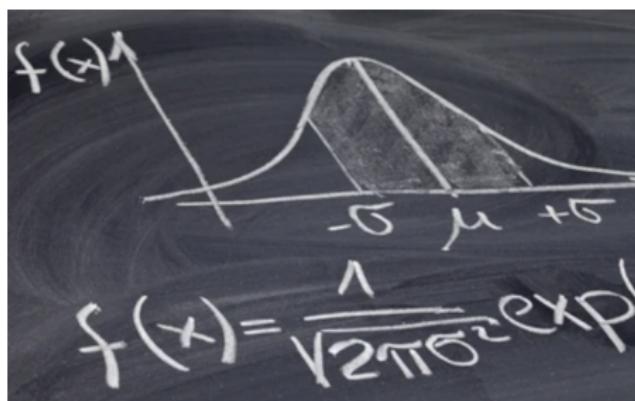
# EL EJEMPLO DE KEPLER



Kepler estudió las complejas tablas de datos compilada por Tycho Brahe y logró extraer de ellas sus famosas y muy compactas leyes del movimiento planetario



# ESTADÍSTICA: CIENCIA PARA EL APRENDIZAJE



La estadística relaciona los datos con cuestiones (preguntas) específicas de interés. Para ello, estudia y crea métodos para: recopilación de información relevante, extracción de respuestas a las preguntas y su respaldo por los datos y finalmente, métodos de resumen y presentación de la información.



# ACERCA DE LOS DATOS

Los datos casi siempre contienen incertidumbre, esta puede surgir de la selección de los elementos a medir, o de la variabilidad de los procesos de medición.

La extracción de conclusiones generales de los datos (inferencia) es la base para conocimiento cada vez mayor, y constituye los cimientos de la investigación científica.

La estadística provee métodos y herramientas para poder hacer inferencias e incluso predicciones a pesar de la incertidumbre en los datos. Los métodos utilizados para el análisis dependen de la forma en que se recopilaron los datos. Es de vital importancia que exista un modelo probabilístico que explique la incertidumbre en los datos.



# ¿QUÉ ES LA CIENCIA DE DATOS I?

- Es un campo interdisciplinario que utiliza métodos científicos, procesos, algoritmos y sistemas para **extraer conocimiento e información de datos ruidosos, estructurados y no estructurados.**
- La ciencia de datos está relacionada con la minería de datos, el aprendizaje automático y el big data.



# ¿QUÉ ES LA CIENCIA DE DATOS II?



# ¿QUÉ ES LA CIENCIA DE DATOS III?

Es un concepto que unifica estadística, análisis de datos, ciencias de la información y de la computación con el fin de comprender fenómenos a partir de datos obtenidos de ellos.



# INTELIGENCIA ARTIFICIAL (IA)

Es una rama de las ciencias de la computación cuyo objetivo consiste en lograr que los computadores tomen decisiones emulando el comportamiento humano.



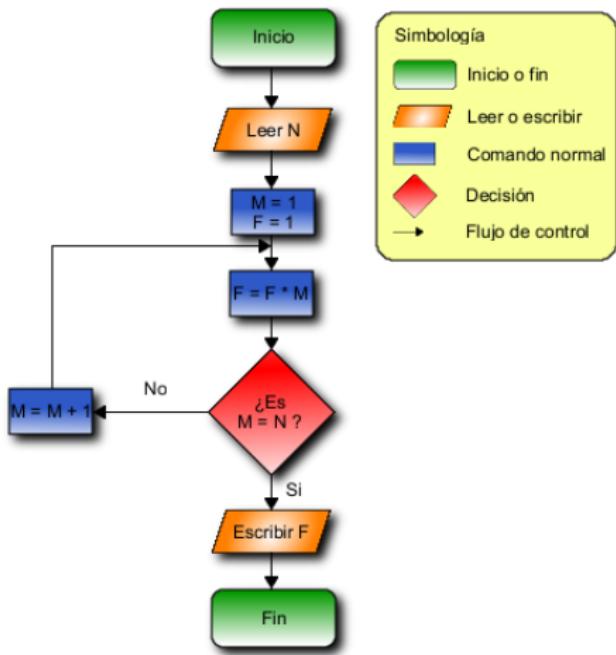
IA

## OBSERVACIÓN

A diferencia de una decisión cuantitativa, como decidir que número es mayor,  $3/4$  o  $0,333$ , mirar una fotografía y decidir si en ella aparece una montaña, la respuesta a la segunda pregunta, viene de la experiencia, se reconoce a la montaña no como un resultado de la enseñanza algorítmica escolar, sino como la conjunción de las experiencias del pasado contemplando paisajes. **Aprendizaje basado en los datos.**



# PROGRAMACIÓN IMPERATIVA



- **Algoritmo.** Secuencia de instrucciones precisas.
- La ejecución transforma la **entrada** al algoritmo en la **salida** (solución) esperada.



# EXISTEN TAREAS PARA LAS CUALES NO EXISTE ALGORITMO ALGUNO

## DETECCIÓN/CLASIFICACIÓN DE SPAM

- ① **Entrada:** Mensajes de e-mail
  - ② **Salida:** Booleana, un simple si o no nos indicando si un mensaje es o no deseado.
- Clasificación personalísima (la basura de una persona puede ser el tesoro de otra).
  - **Ejercicio previo:** clasificación (labelling) de una muestra de mensajes por parte del usuario
  - El resultado del labelling es utilizado por un computador para “aprender” cuales mensajes constituyen spam y cuales no.



# IA EJEMPLOS

Son muchas las veces en que se toman decisiones sobre la base de datos insuficientes, dudosos y hasta inconsistentes, en esos casos, el conocimiento previo (experiencia) es la herramienta de elección.

- Diagnóstico médico
- Reconocimiento de caracteres escritos (OCR) y de rostros



# APRENDIZAJE ESTADÍSTICO MACHINE LEARNING, ML

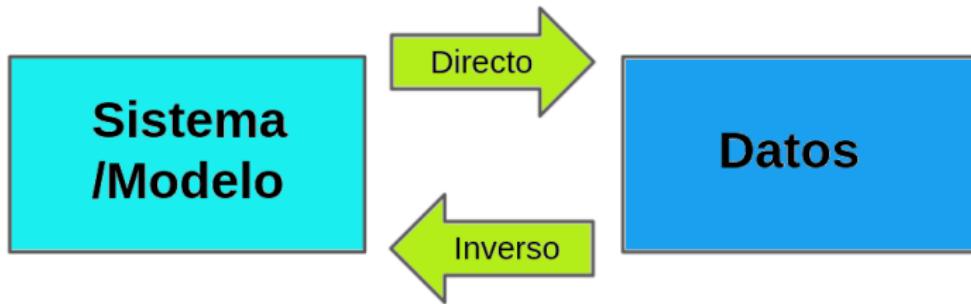
## DEFINICIÓN

*El Aprendizaje Estadístico es un método de análisis de datos que automatiza la construcción de modelos analíticos. Es una rama de la inteligencia artificial basada en la idea de que los sistemas pueden aprender de los datos, identificar patrones y tomar decisiones con una mínima intervención humana.*



# PROBLEMAS DIRECTOS E INVERSOS

$$\mathbf{d} = \mathcal{F}(\mathbf{m})$$



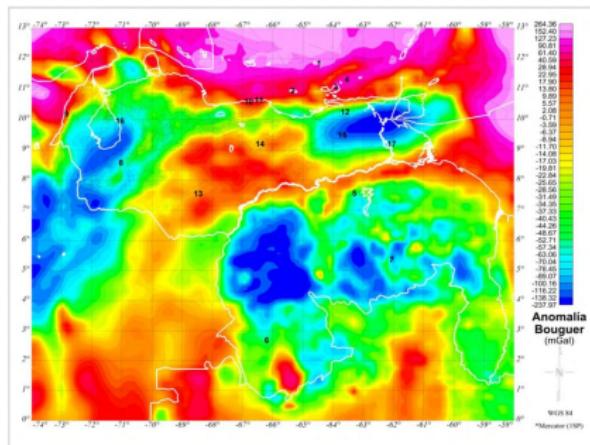
$$\mathbf{d}^{(obs)} = \mathcal{F}(\mathbf{m})$$



# PROBLEMA DE APRENDIZAJE

## PROBLEMA DE AJUSTE

### PROBLEMA INVERSO



- **Datos** Valores de aceleración de gravedad en una región.
- **Modelo** Densidad de masa en el subsuelo.



# MODELO PROBLEMA DIRECTO

## Problema Directo

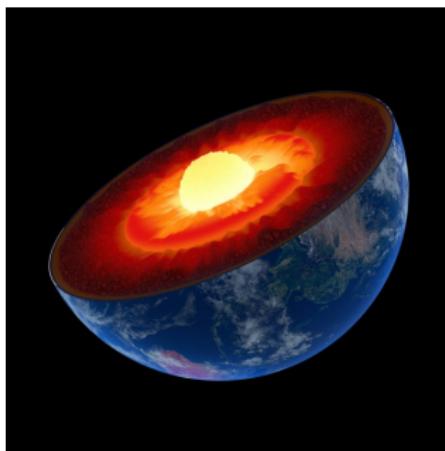
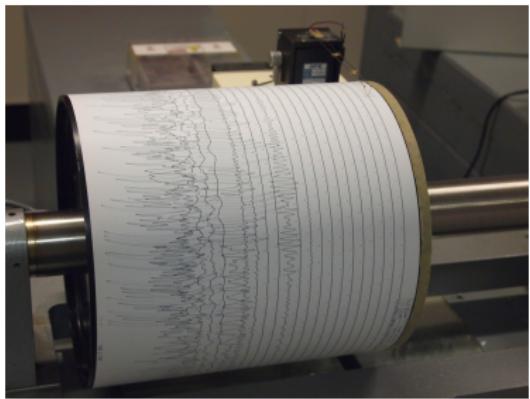
Conocida la distribución de masa en una región del espacio,  
encuentre el campo gravitacional en la superficie de la región

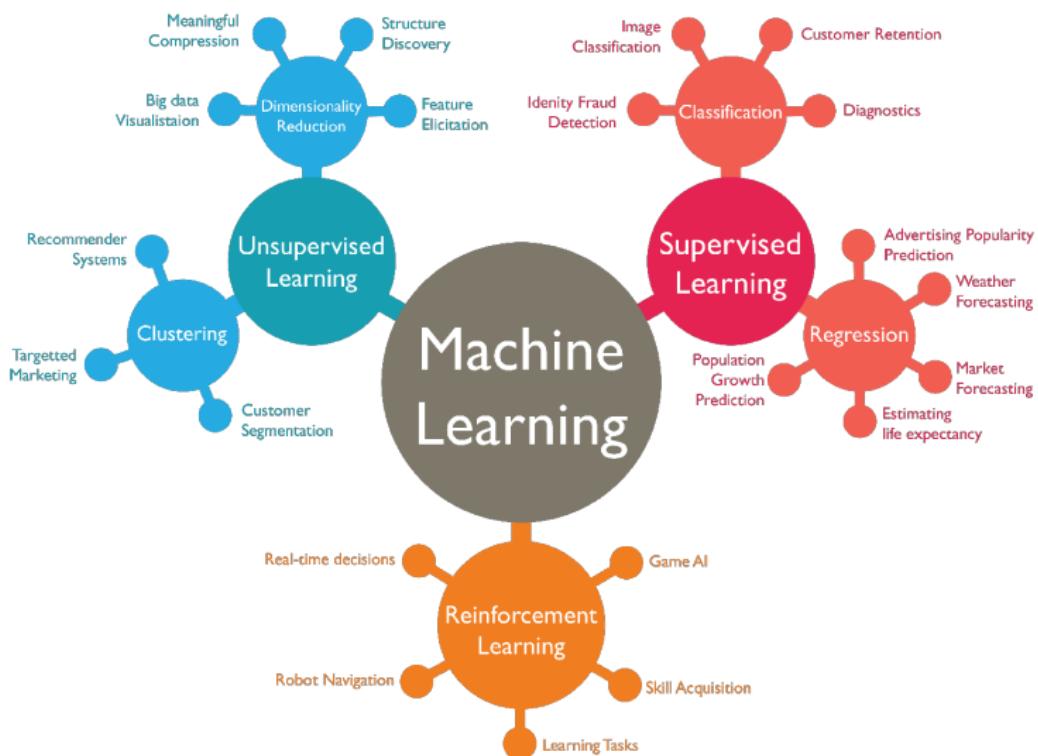
$$\mathcal{G}(\mathbf{x}) = G \int \frac{\rho(\mathbf{x}') (\mathbf{x} - \mathbf{x}') d^3x'}{|\mathbf{x} - \mathbf{x}'|}$$



# REGRESIÓN≈PROBLEMA INVERSO

$$\mathbf{d}^{(obs)} = \mathbf{F}(\mathbf{m})$$





# EL PROBLEMA DE LENGUAJE

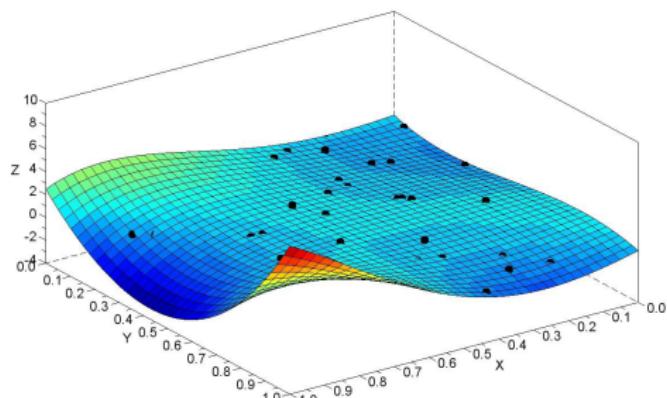
Por razones tanto históricas como de inter y transdisciplinariedad, en ML se consigue una enorme cantidad de vocabulario ad hoc que complica las cosas

- Entrenamiento (Training)
- Conjuntos de prueba y entrenamiento (Training & Testing sets)
- Regularización
- Métodos de Gradiente
- Batch
- HyperParameters



# REGRESIÓN≈PROBLEMA INVERSO

$$z_i = z_0 + a_i x_i + b_i y_i + c_i x_i^2 + d_i y_i^2 + e_{ij} x_i y_j + [\dots] + \epsilon_i$$



- $\mathbf{d} = \mathbf{F}(\mathbf{m})$ .
- $\mathbf{y} = h(\Theta|\mathbf{x})$



# PREREQUISITOS

- Álgebra lineal, cálculo en varias variables, probabilidades, estadística.
- Elementos de programación
- Entendimiento de los datos



# PROBLEMA DE APRENDIZAJE I

- **Datos**  $N$  pares de “mediciones” ( $\mathbf{y}^i, \mathbf{x}^i$ ), de las variables objetivo y sus predictores.
- **Modelo/Hipótesis**

$$\boxed{\mathbf{y}^i = h(\Theta | \mathbf{x}^i)}$$

donde  $\Theta$ , el vector de pesos es el objeto de aprendizaje

- Las dimensionalidades de las variables objetivo, los predictores y los pesos son, en general, distintas.



# PROBLEMA DE APRENDIZAJE II

## PROBLEMA

*Dados los datos (variables objetivo  $\mathbf{Y}$  y predictores/atributos  $\mathbf{X}$ ) y la hipótesis/modelo*

$$\mathbf{Y} = \mathbf{H}(\boldsymbol{\Theta} | \mathbf{X})$$

*calcular (aprender) los pesos*



# SOLUCIÓN AL PROBLEMA DE APRENDIZAJE

- ① Construcción de una función de *pérdida* (loss)

$$L[\mathbf{Y}; \mathbf{H}(\Theta|\mathbf{X})]$$

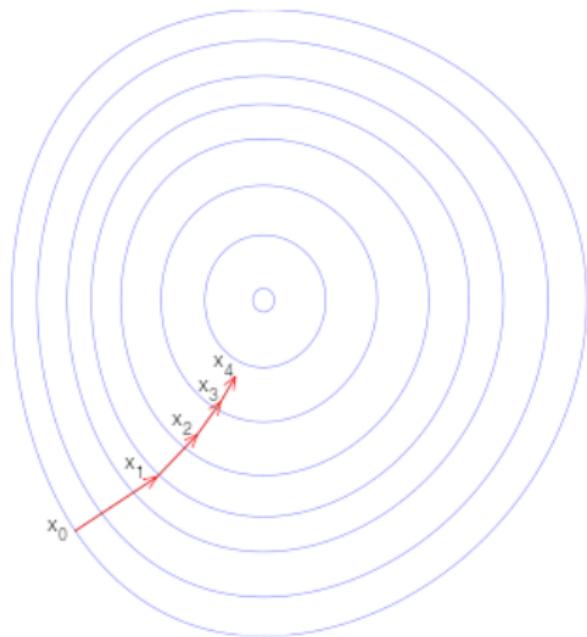
- ② Los pesos *aprendidos*  $\hat{\Theta}$  corresponden a la solución del problema de optimización

$$\hat{\Theta} = \operatorname{argmin}_{\Theta} \{L[\mathbf{Y}; \mathbf{H}(\Theta|\mathbf{X})]\}$$

- ③ La solución estándar del problema de optimización es a través del *algoritmo de descenso por gradiente*



# DESCENSO POR GRADIENTE I



- El método de descenso por gradiente es un **algoritmo** de minimización **iterativo**
- El algoritmo utiliza la propiedad geométrica fundamental del gradiente.



# DESCENSO POR GRADIENTE II

- ① El algoritmo comienza dando una semilla (valor inicial)  $\Theta^{(0)}$
- ② A partir de la semilla se itera (el superíndice k indica la iteración) para conseguir nuevos valores de los parámetros

$$\Theta^{(k+1)} = \Theta^{(k)} - \mu \nabla_{\Theta^T} L(\Theta^{(k)})$$

- ③ La iteración se detiene con algún criterio. [Un criterio típico consiste en la estabilización de la función de costo].



# PSEUDOCÓDIGO

---

## Descenso por Gradiente

**procedure** GRADIENT DESCENT

**Input**  $\Theta^{(0)}$ : Semilla

$\Theta^{(k)} = \Theta^{(0)}$

**While**  $epsilon \geq criterio$

$L^{(k)} = L(\Theta^{(k)})$

**Paso de Gradiente:**

$$\Theta^{(k+1)} = \Theta^{(k)} - \mu \nabla_{\Theta^T} L(\Theta^{(k)})$$

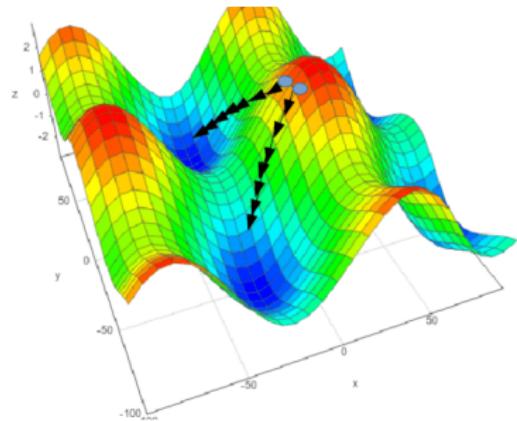
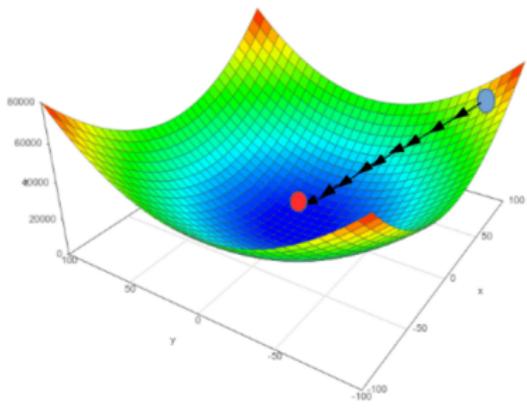
$$L^{(k+1)} = L(\Theta^{(k+1)})$$

$$\epsilon = |L^{(k+1)} - L^{(k)}|$$

**Output:** return  $\Theta^{(k+1)}$



## DESCENSO POR GRADIENTE IV



- Evidentemente hay montones de tecnicismos que no estamos tocando en esta exposición, entre otros.
  - Escogencia de  $L$ .
  - Peprocesamiento de los datos.
  - Validación de resultados.
  - Predicciones
- Recordemos que ML es una subdisciplina de la Ciencia de Datos, otras subdisciplinas, como Big Data, por ejemplo, solo han sido mencionadas



## ● Lenguajes

- ① R
- ② Python

## ● Bibliotecas

- ① caret, R
- ② Sklearn, Python
- ③ TensorFlow, Python Neural Networks
- ④ Pytorch, Python Neural Networks
- ⑤ Keras, Python Framework para TensorFlow

## ● Recursos Cloud

- ① R Studio Cloud
- ② Colab
- ③ Kaggle

