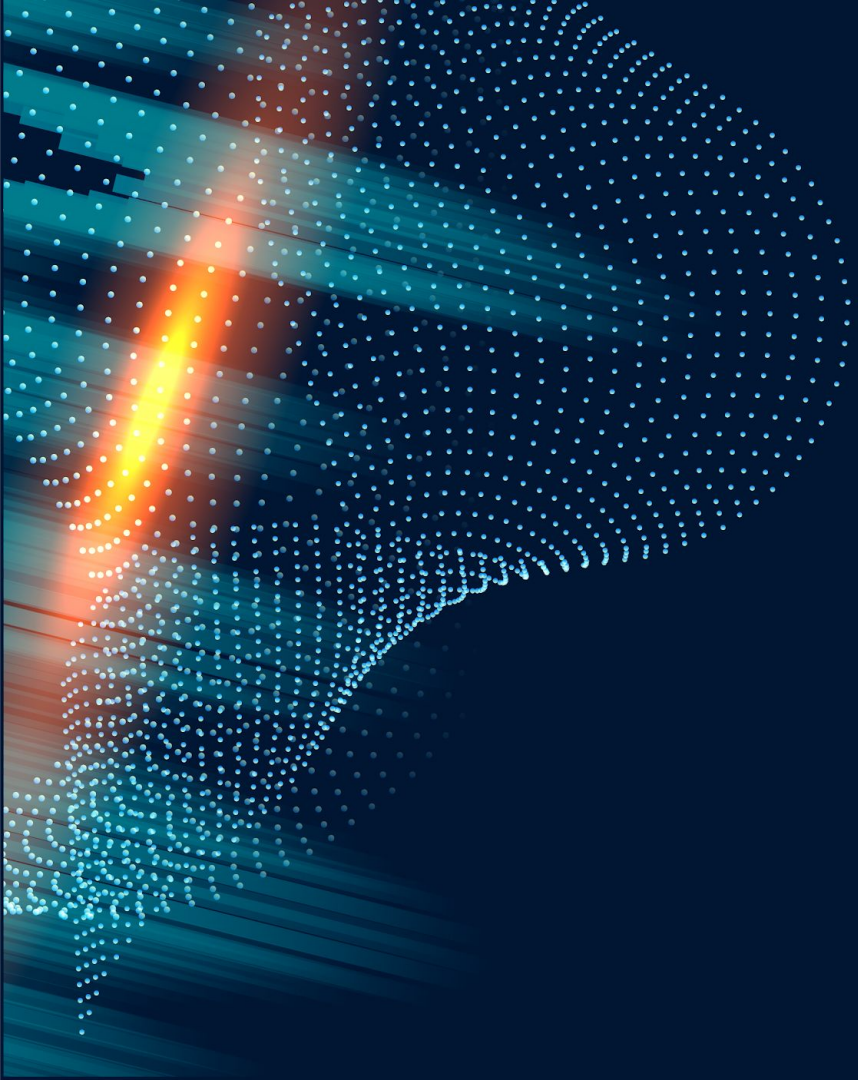
Machine Learning 101

Aquí inicia tu viaje por la ciencia de datos

José Ángel Medina

2 de septiembre de 2021



José Ángel Medina

CDO PiLab y Ridery

PiLAB Ridery

CONTENIDO

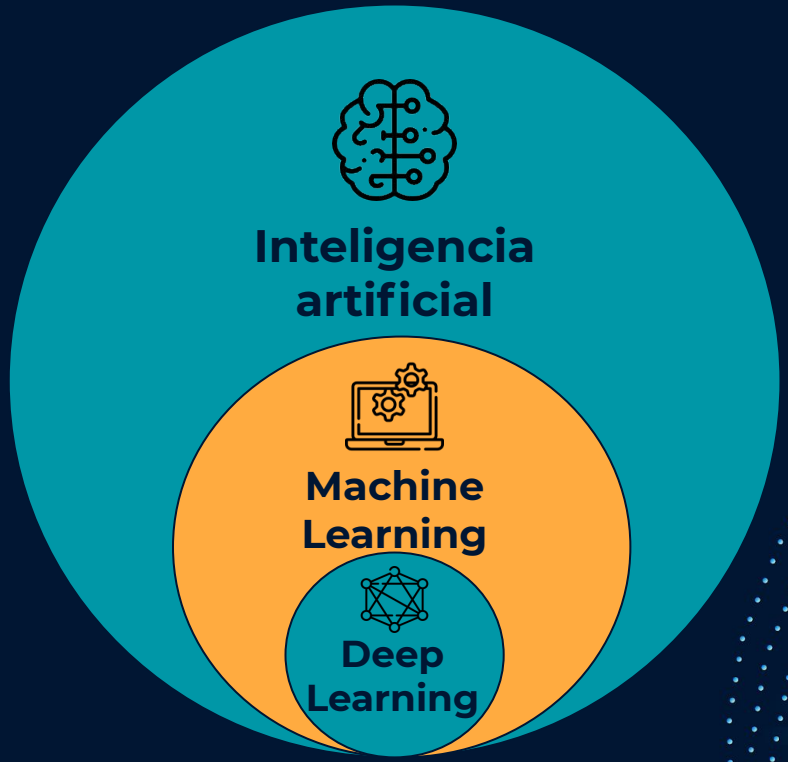
1. ¿Qué es Machine Learning?
2. ¿Para qué usar Machine Learning?
3. Algunas aplicaciones
4. Tipos de Machine Learning
5. Algoritmos supervisados y no supervisados comunes
6. Retos en la construcción de modelos
7. Training and testing
8. Modelos en la práctica
9. Línea de tiempo de un proyecto de ML y ciclo de vida de un modelo
10. ¿Cómo elegir el modelo adecuado?
11. ¿Dónde aprender más ML?

¿Qué es Machine Learning?

Cuando las máquinas aprenden...

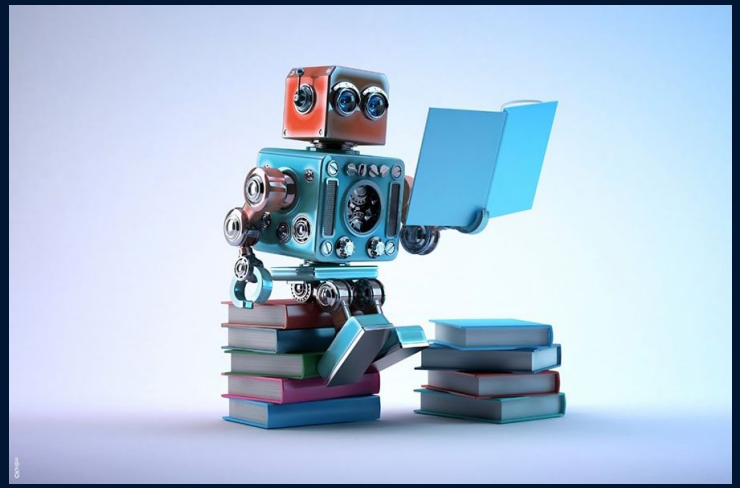
(Inserte teoría conspirativa aquí)

¿Qué es Machine Learning?



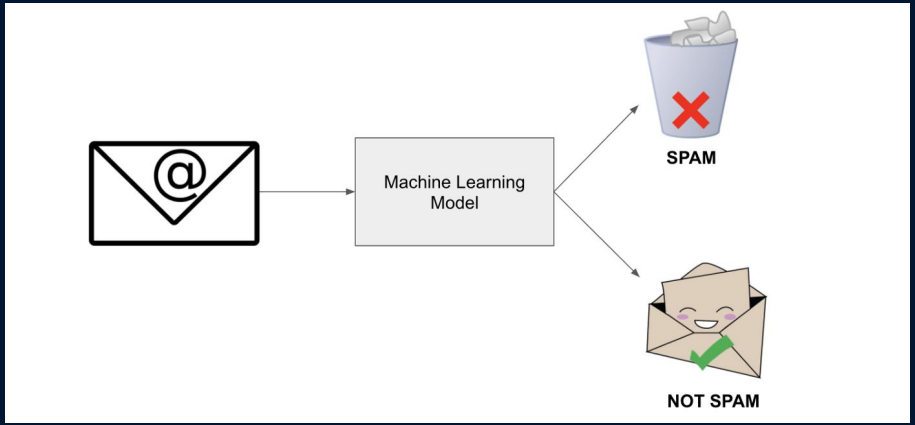
¿Qué es Machine Learning?

“ML es el arte y ciencia de programar computadoras para que puedan aprender de data.” - Aurelien Geron



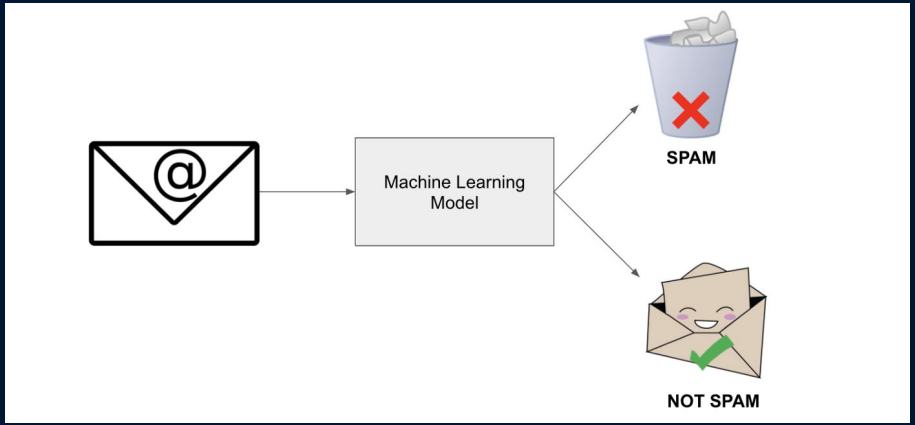
¿Qué es Machine Learning?

El filtro de spam es uno de los primeros programas que ha simplificado la vida usando ML



¿Qué es Machine Learning?

Este programa aprende cuando manualmente marcamos un correo como spam



¿Para qué usar ML?

¿Para qué usar ML?

Mucho que ajustar

Problemas que requieren muchos ajustes o adaptación frecuente del algoritmo

¿Para qué usar ML?

Mucho que ajustar

Problemas que requieren muchos ajustes o adaptación frecuente del algoritmo

Muy complicado

Problemas como reconocimiento del habla que son muy complejos para *"codear"*

¿Para qué usar ML?

Mucho que ajustar

Problemas que requieren muchos ajustes o adaptación frecuente del algoritmo

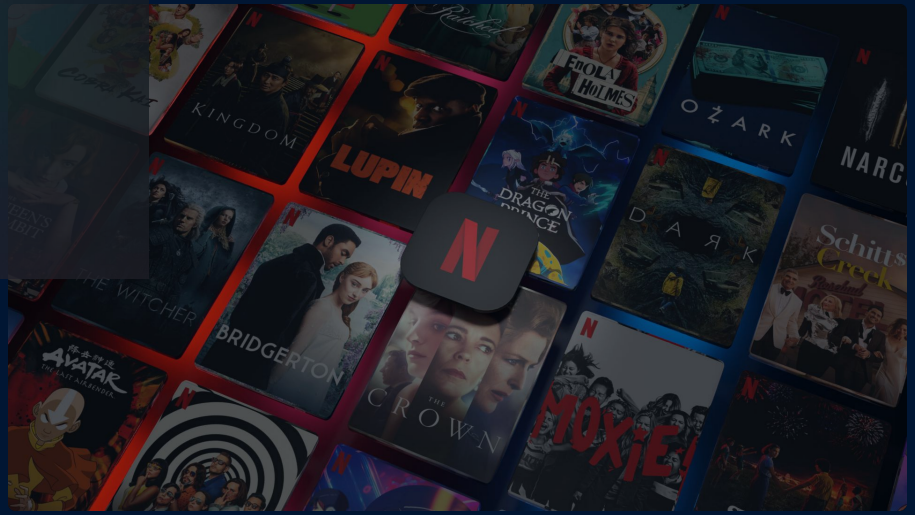
Muy complicado

Problemas como reconocimiento del habla que son muy complejos para *"codear"*

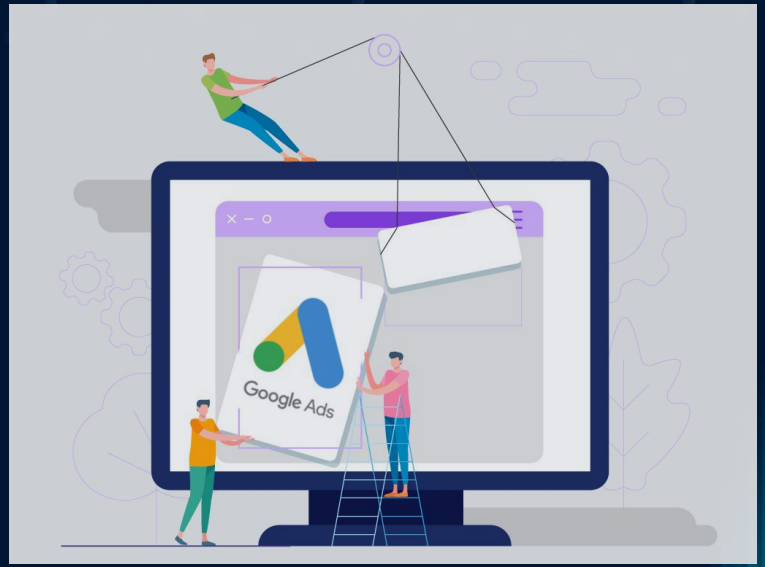
Eso no lo sabía

Para identificar patrones complejos en cantidades masivas de data

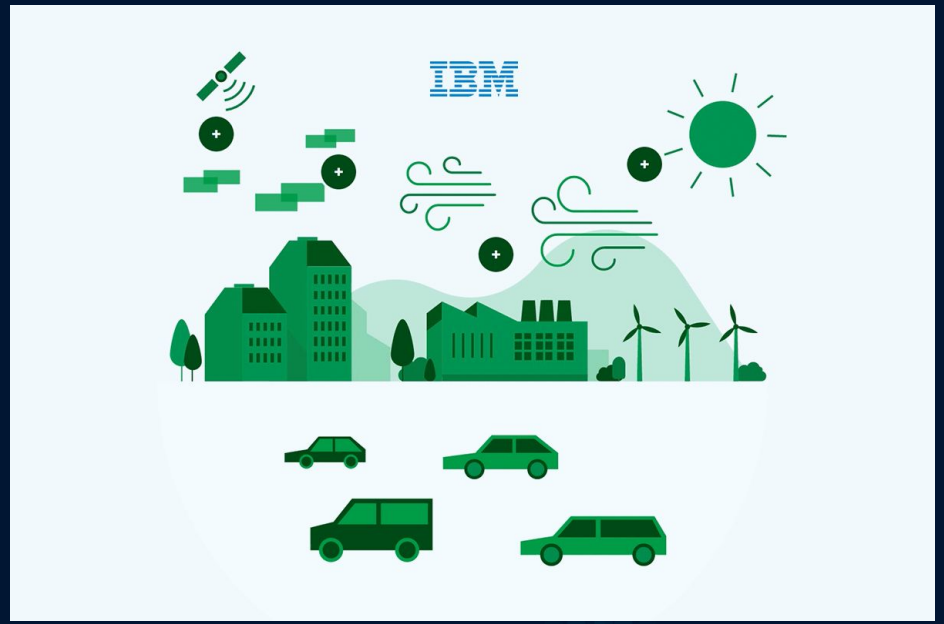
Algunas aplicaciones



Algunas aplicaciones



Un caso de estudio: Green Horizon



- Inició en 2014
- Objetivo: mejorar la calidad del aire en China
- Aprovecha ambiental data pública
- Utiliza ML para predecir la contaminación del aire
- Integra tecnologías como IoT
- Para 2017 había disminuido las partículas peligrosas en el aire en 25%

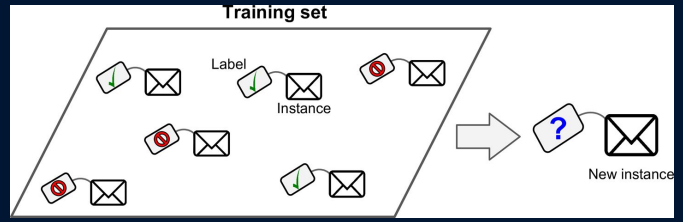
Tipos de Machine Learning

Supervisado
Vs
No supervisado

Batch
vs
Online

Instancias
vs
Modelo

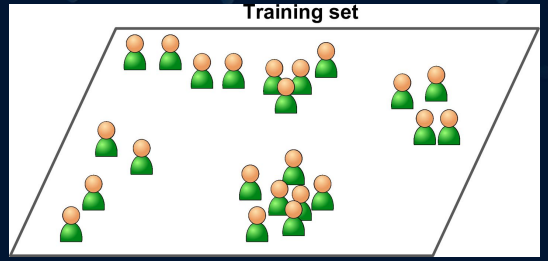
Aprendizaje supervisado vs no supervisado



Supervisado: Aprende de sets de datos que han sido etiquetados por humanos

Tareas comunes:

- **Regresión**
- **Clasificación**



No supervisado: Descubre patrones en sets que no han sido etiquetados

Tareas comunes:

- **Clustering**
- **Detección de anomalías**
- **Reducción dimensional**
- **Regla de asociación**

Imagenes extraidas de “Hands-On Machine Learning with Scikit-Learn, Keras and Tensor Flow” por Aurelien Geron publicado por O’Reilly. El uso de estas imágenes en esta presentación es educativo y no está asociado a lucro alguno. Todos los derechos a sus dueños

Algoritmos supervisados comunes

- Regresión lineal
- Regresión logística
- Support Vector Machines (SVM)
- Decision Trees y Random Forests
- Redes neuronales

Algoritmos no-supervisados comunes

- k-Means
- DBSCAN
- Isolation Forests
- Principal Component Analysis
- Kernel PCA
- Apriori
- Eclat

¡Peligro!

Crear un modelo no es tan fácil como escribir el código y ya

La data es más importante que el algoritmo

Si tienes muy poca data, el algoritmo no va a funcionar bien

La data debe ser representativa

La data de entrenamiento de representar a la data de producción

Garbage in, garbage out

Hay que limpiar la data antes de usarla y elegir las variables adecuadamente

Beware of Overfitting!

and underfitting too

Beware of Overfitting!

and underfitting too

Hyperparameter: parámetros del algoritmo (no del modelo) que se mantienen fijos a pesar del aprendizaje o los datos. Se usan para regularizar el modelo

Training and Testing

Para evitar que un modelo con problemas entre a producción, lo probamos usando data para la que ya conocemos el resultado

Dataset

Training set

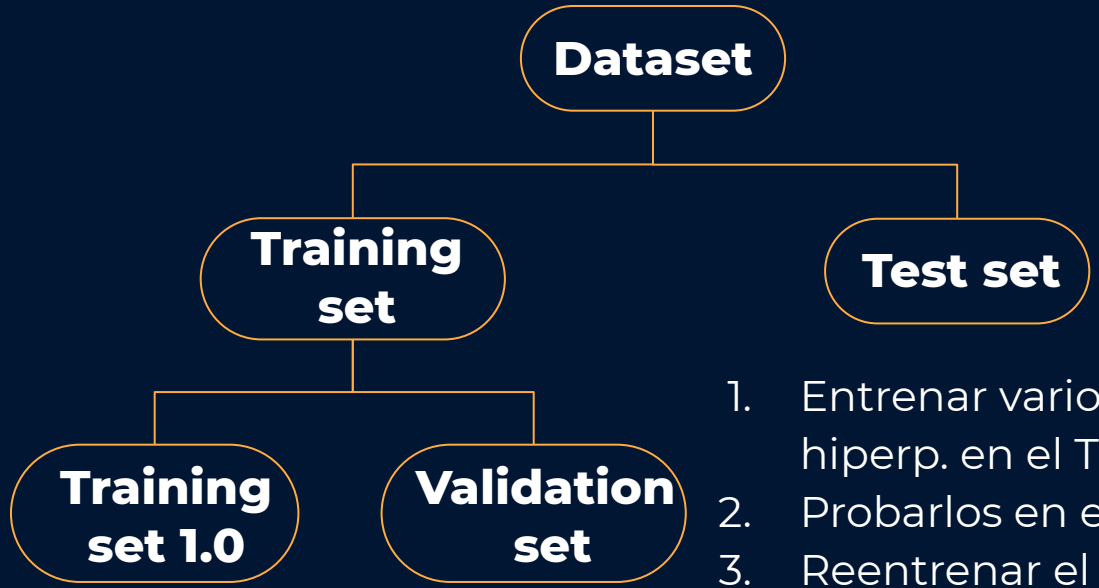
80% del dataset

Test set

20% del dataset

La división debe realizarse al azar para evitar biases

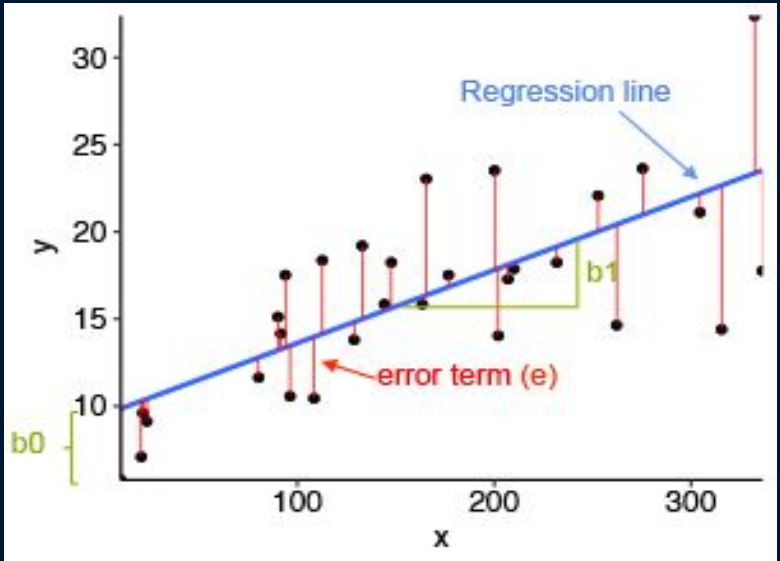
Validación para selección de modelos



1. Entrenar varios modelos con varios hiperp. en el Training Set 1.0
2. Probarlos en el Validation Set
3. Reentrenar el mejor en el Training set total (1.0+VS)
4. Probar con el Test Set

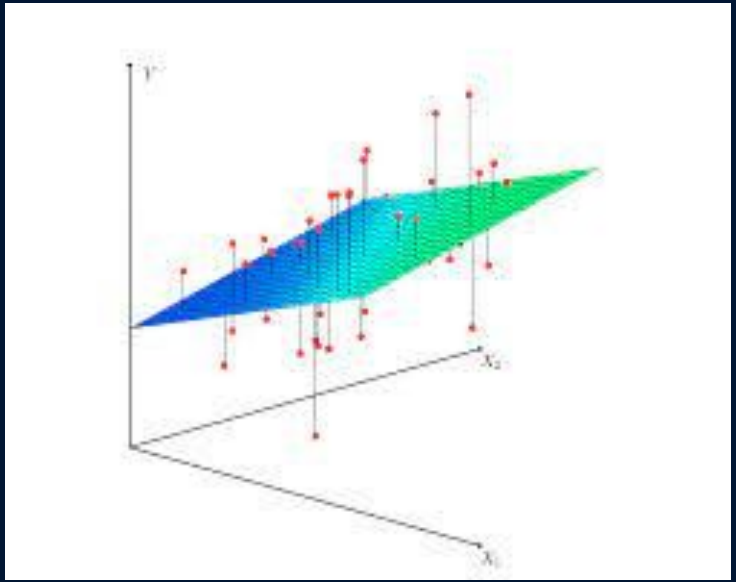
scikit
learn
pandas

Regresión lineal



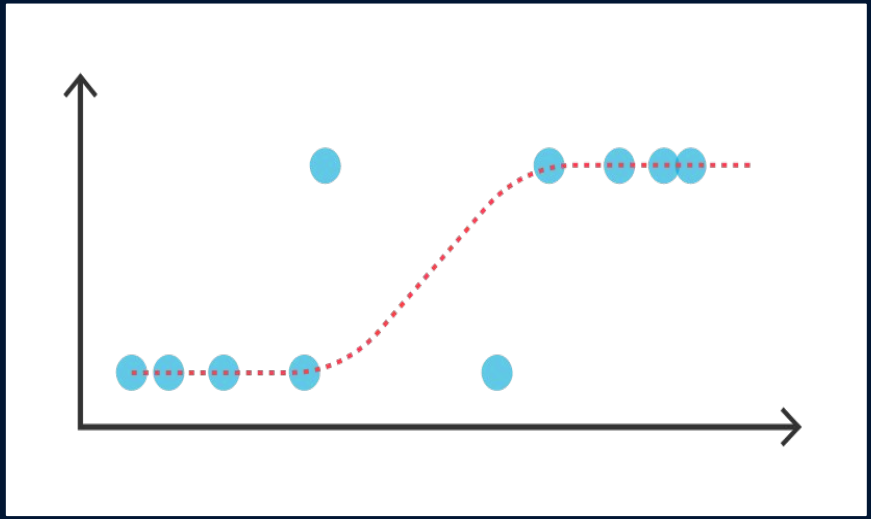
y^ = b0 + b1*x

Regresión lineal



$$y\hat{} = b0 + b1*x1 + b2*x2 + \ldots + bn*xn$$

Regresión logística



$$f(x) = (1 + e^{-y})^{-1}$$

$$y = b0 + b1*x1 + \ldots + bn*xn$$

Línea de tiempo de un proyecto de ML

1 Entender el problema

2 Establecer una medida de desempeño

3 Conseguir la data

4 Preparar la data 1

Línea de tiempo de un proyecto de ML

5

Armar el training y test sets

Analizar el training set para descubrir

6

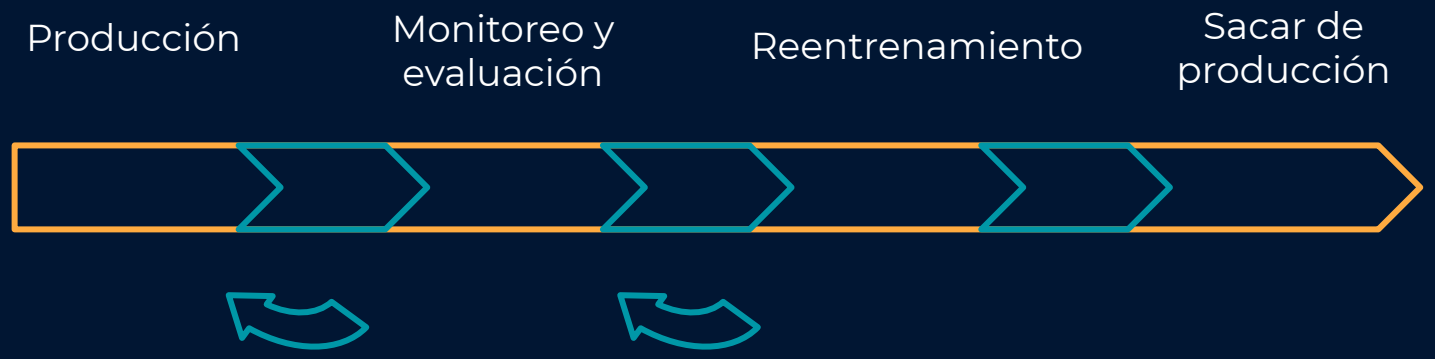
7

Preparar la data 2

Machine Learning away!

8

Ciclo de vida de un modelo



¿Cómo elegir el algoritmo adecuado?

Consideremos:

- ¿Cuál es el problema a resolver?
- ¿Qué tipo de problema es?
  - ¿Regresión o clasificación?
- ¿Cómo viene la data?
  - ¿Etiquetada o no etiquetada? ¿En bache o en línea?
- ¿Qué dice mi análisis?
- ¿Es más importante dar resultados rápidos o precisos?
- ¿Estoy teniendo overfitting por más que ajuste los hp?
- Una vez en producción, ¿dónde correrá el modelo?

¿Dónde aprender ML?

www.datasmarts.net/es/

- Gratis
- Contenido en español
- Especializado en Computer Vision

www.superdatascience.com/challenge

- Gran variedad de cursos
- Tiene una hoja de ruta Consigue trabajo en 99 días
- $35 / m o $23 / m
- Podcast Free

www.coursera.org

- Gran variedad de cursos
- Certificaciones por Google y Universidades
- Precios variables

¿Dónde aprender ML?

YouTube

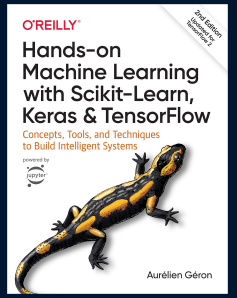
- Ken Jee
- Tina Huang
- Luke Barousse
- Krish Naik
- Joma Tech

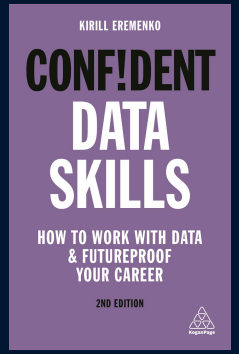
kaggle

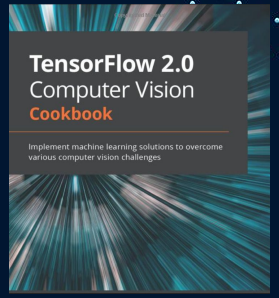
www.kaggle.com

- Comunidad
- Muchos, muchos datasets free
- Challenges
- Algunos cursos

¿Dónde aprender ML?







Libros, obvio:

- Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems: Géron, Aurélien: 9781492032649: Amazon.com: Books
- Amazon.com: Confident Data Skills: How to Work with Data and Futureproof Your Career (Confident Series) eBook : Eremenko, Kirill: Kindle Store
- TensorFlow 2.0 Computer Vision Cookbook: Implement machine learning solutions to overcome various computer vision challenges: Martinez, Jesús: 9781838829131: Amazon.com: Books

“Una computadora merecerá ser llamada inteligente si logra hacerle creer a una persona que ella es humana.”

— Alan Turing

DANKE!

¿Preguntas?
joseangelmc.ds@gmail.com

joseangelmc

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik.

slidesgo