

Introducción a Machine Learning

Taller invitado
Gestión de Operaciones
Escuela de Ingeniería Industrial - ULACIT



Colombia



Texas, USA



Costa Rica

Headquarters



300+ 

full time

GAPSTERs en LATAM



Nuestras Áreas



Analytics



Mobile



Cloud



**QA/
QA Automation**

Contenido

Introducción a Machine Learning
Estrategias de manufactura
ULACIT

01.

Conceptos

02.

Etapas del
proceso

03.

Consideraciones
éticas

04.

Aplicaciones
en Ingeniería
Industria;

05.

Aplicación

06.

Aplicación

Machine Learning

Datos

Unidades de información asociadas a un fenómeno observado.

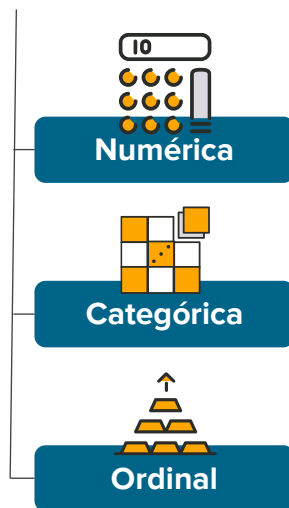
¿Por qué es cada vez más importante?

- Crecimiento de plataformas digitales.
- Digitalización de servicios.
- Nuevas metodologías de análisis desbloquean nuevos potenciales.
- Más datos => Más responsabilidad.

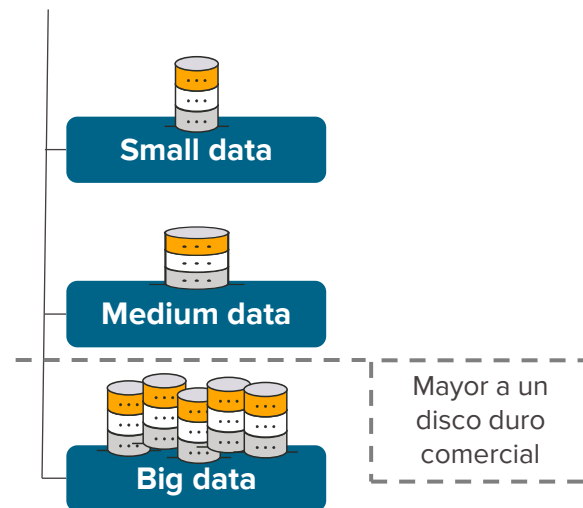
Cómo se almacena la información



Qué tipo de información contiene

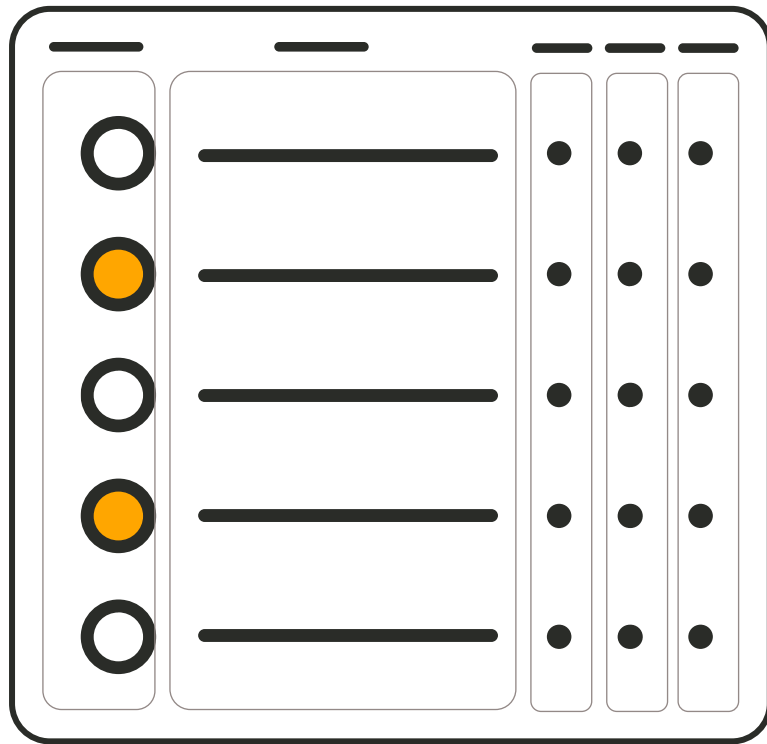


De qué tamaño es el conjunto de datos



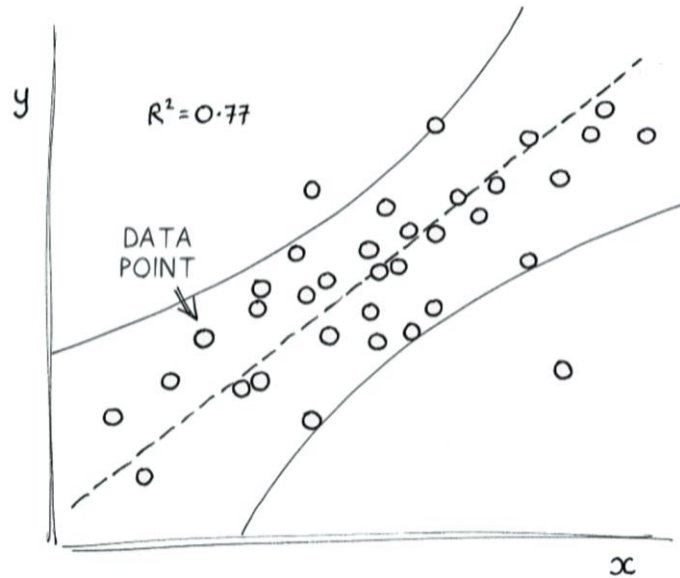
Datos estructurados

- Tablas como colecciones de vectores
- Mucha algebra lineal



Modelación Matemática

Construcciones que buscan aproximar la realidad a través de expresiones matemáticas.



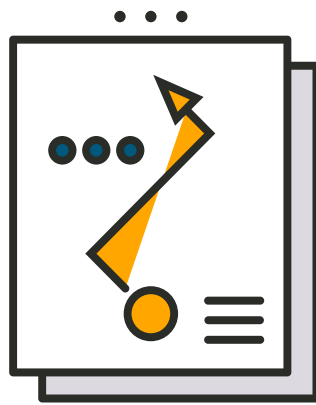
$$C = S \cdot N(d_1) - X \cdot e^{-r \cdot T} \cdot N(d_2)$$

Donde d_1 y d_2 son tal que:

$$d_1 = \frac{\ln \frac{S}{X} + \left[r + \frac{\sigma^2}{2} \right] \cdot T}{\sigma \cdot \sqrt{T}}$$

$$d_2 = \frac{\ln \frac{S}{X} + \left[r - \frac{\sigma^2}{2} \right] \cdot T}{\sigma \cdot \sqrt{T}} = d_1 - \sigma \cdot \sqrt{T}$$

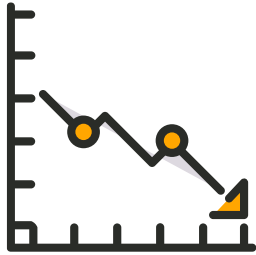
Machine Learning



Campo de estudio que construye sistemas capaces de aprender, en lugar de ser explícitamente programados.

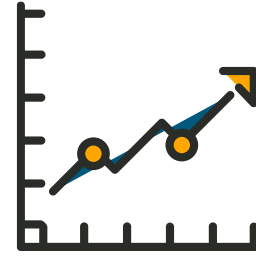
- Los modelos de machine learning mejoran automáticamente a través de experiencia (exposición a datos).
- Los procesos de entrenamiento siguen reglas conocidas. **¡No son ni deben usarse como cajas mágicas!**

Machine learning es optimización



Minimizar

Errores en valores
Calsificaciones incorrectas



Maximizar

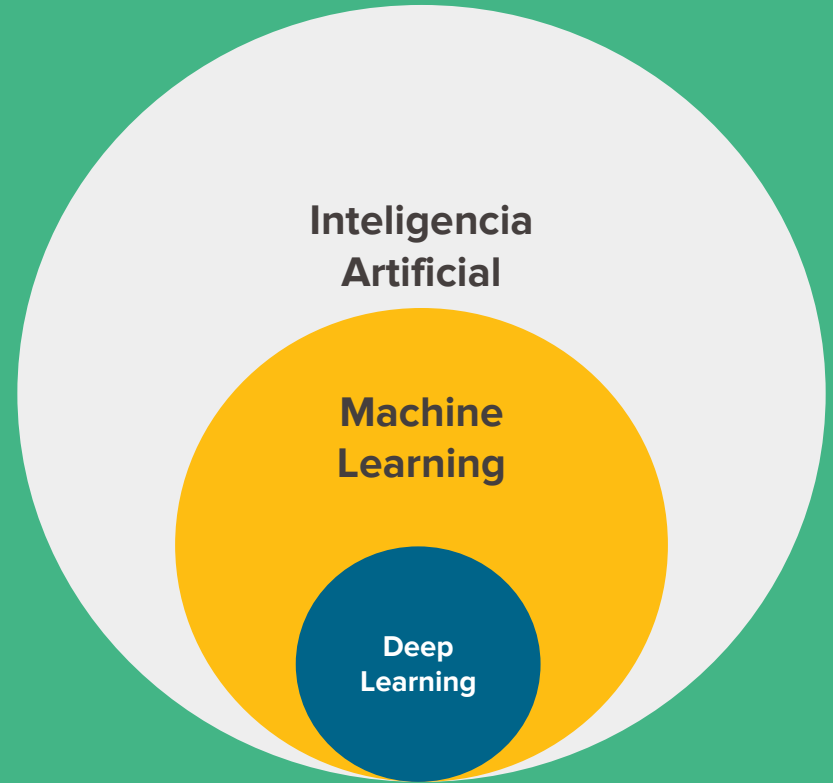
Diferencias entre
categorías
Cantidad de información
conservada

Quiénes hacen Machine Learning?



- Científicos y científicas de datos.
- Analistas de datos
- Profesionales de inteligencia de negocios
- Especialistas del área de aplicación*

Machine Learning? Deep Learning? Inteligencia Artificial?



Cómo se hace ML?

Identificar la pregunta

¿Existe una variable que
responda esa pregunta?
¿Podría crearla?



Supervisado

¿De qué tipo es la variable
objetivo?

- Numérica
- Categórica
- Ordinal

Sí

No



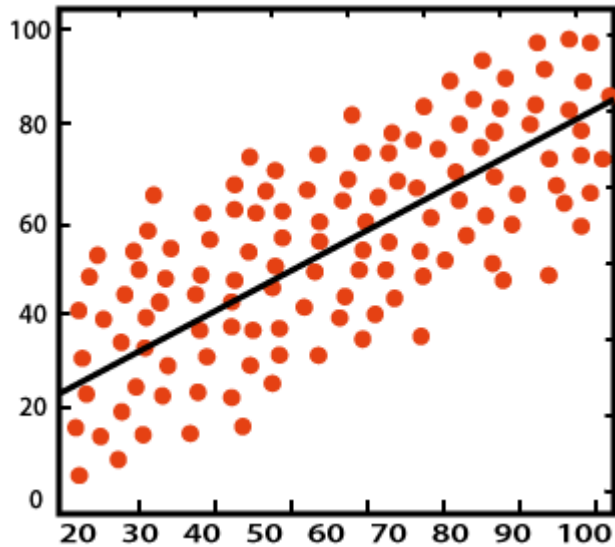
No supervisado

¿Qué quiero entender mejor
en mis datos?

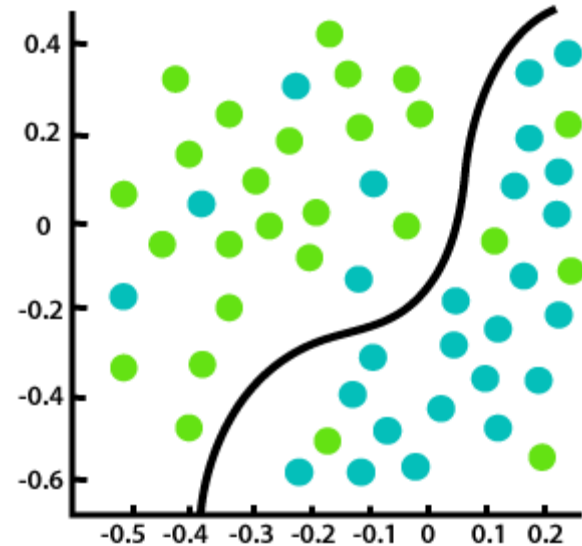
- Cómo se relacionan las
observaciones
- Cómo se relacionan las
características

Aprendizaje supervisado

Regresión

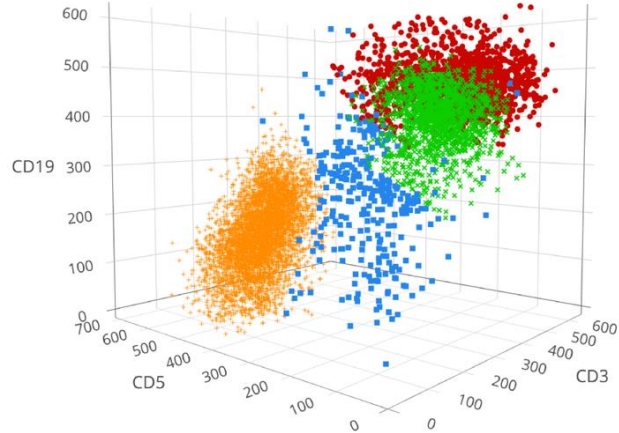


Clasificación

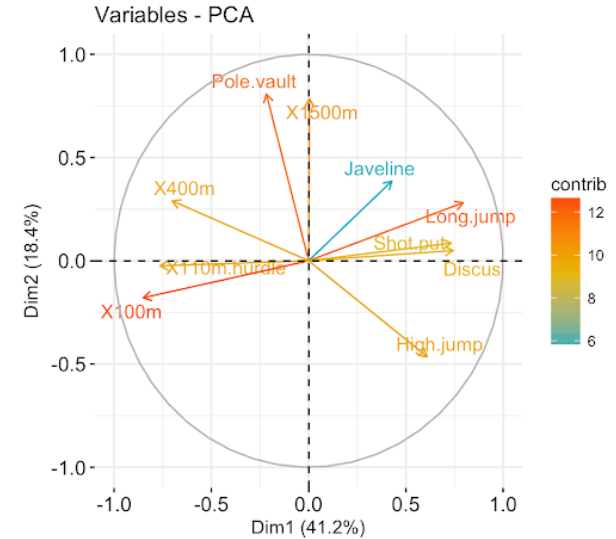


Aprendizaje no supervisado

Segmentación



Reducción de dimensiones



Explorar los datos

1. Entender el contenido.
2. Revisar la calidad de ese contenido (datos faltantes, valores extremos, etc.)
3. Estudiar el comportamiento de cada variable.
4. Estudiar las relaciones entre distintas variables.

Limpiar los datos

- Eliminar?
- Corregir?
- Rellenar?

<input type="radio"/>		●	●	×
<input checked="" type="radio"/>		●	●	●
<input type="radio"/>		×	●	●
<input checked="" type="radio"/>		●	●	
<input type="radio"/>		●	●	●



Requiere
contexto

Feature engineering

Idear, transformar y crear variables que capturen los comportamientos que el queremos usar dentro del modelo



**Requiere
contexto**

Ajustar el modelo

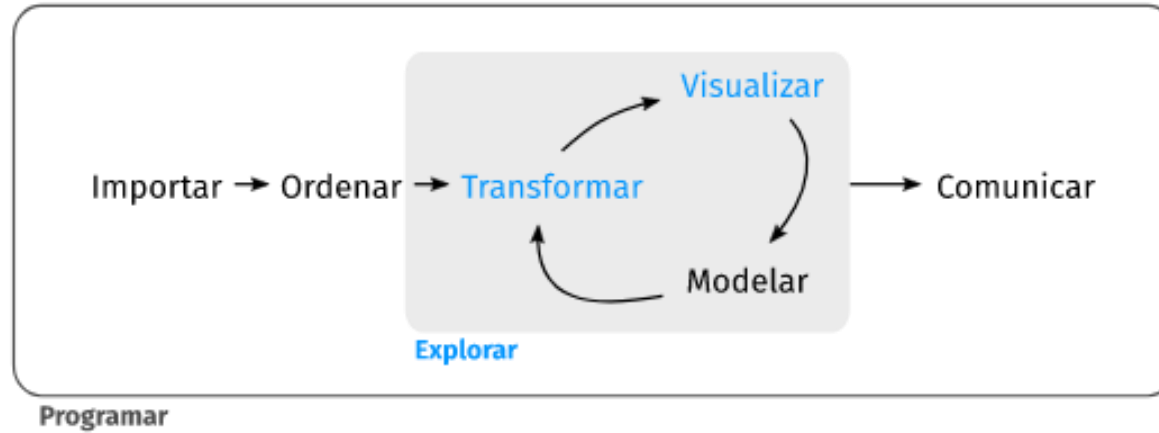
- Elegir el modelo
 - Modelo flexible vs generalizable?
 - Qué tan complejo de interpretar?
- Definir hiper parámetros
 - Hiper parámetros: parámetros del modelo que se deben elegir manualmente para que se pueda dar el proceso de optimización.

Evaluar el modelo

Elegir métricas que se alineen al objetivo del modelo

- Si tengo 97% A y 3% B, puedo tener un 97% de precisión si solo digo que todo es A.
- Si quiero detectar cáncer, es más grave tener falsos negativos que falsos positivos.
- Si quiero predecir ingreso, el error promedio podría ser muy grande por solo unas cuantas observaciones.

Ciclo de trabajo en Machine Learning



Consideraciones éticas

Consideraciones éticas

¿Es adecuado que un sistema de machine learning guíe esas decisiones del todo?

Sobre los datos

¿De dónde vienen?

¿Fueron obtenidos de acuerdo a las respectivas regulaciones?

¿Hay claridad de cómo identificar los sesgos de la data?

Sobre el modelo

¿Entiendo la metodología?

¿Conozco qué está considerando para dar sus resultados?

¿Qué dice sobre el fenómeno que se está estudiando?

¿Es suficientemente preciso para guiar decisiones?

¿Debería hacerse un modelo del todo?

Referentes

Timnit Gebru
@timnitGebru

Margaret Mitchell
@MMitchell_ai

Rachel Thomas
@math_Rachel

<https://ethics.fast.ai/>

Documental “Coded Bias”

Aplicaciones en ingeniería industrial

Análisis de cadenas de suministros

Optimización dinámica de rutas

Tradicionalmente:

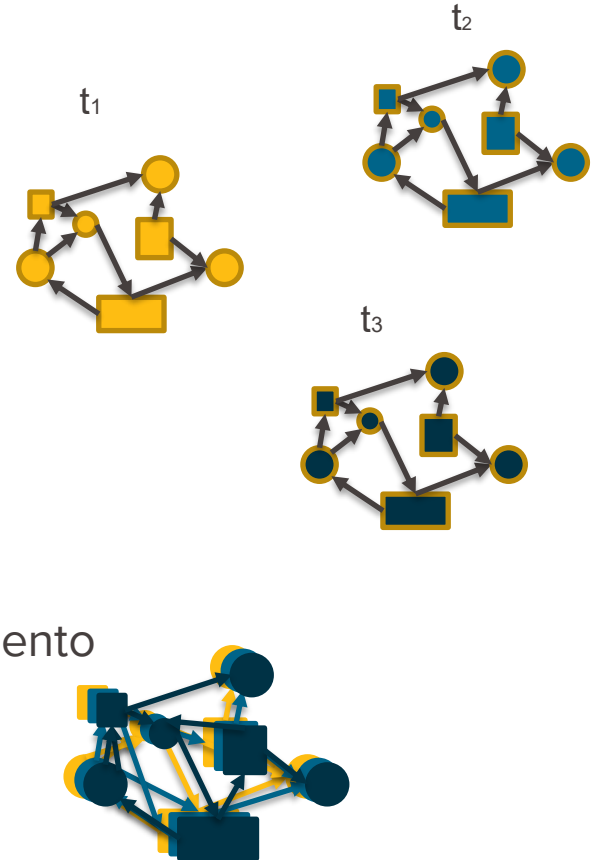
- Recoger muchos datos.
- Construir la mejor ruta posible para esos datos.
- 🖱️

Online learning

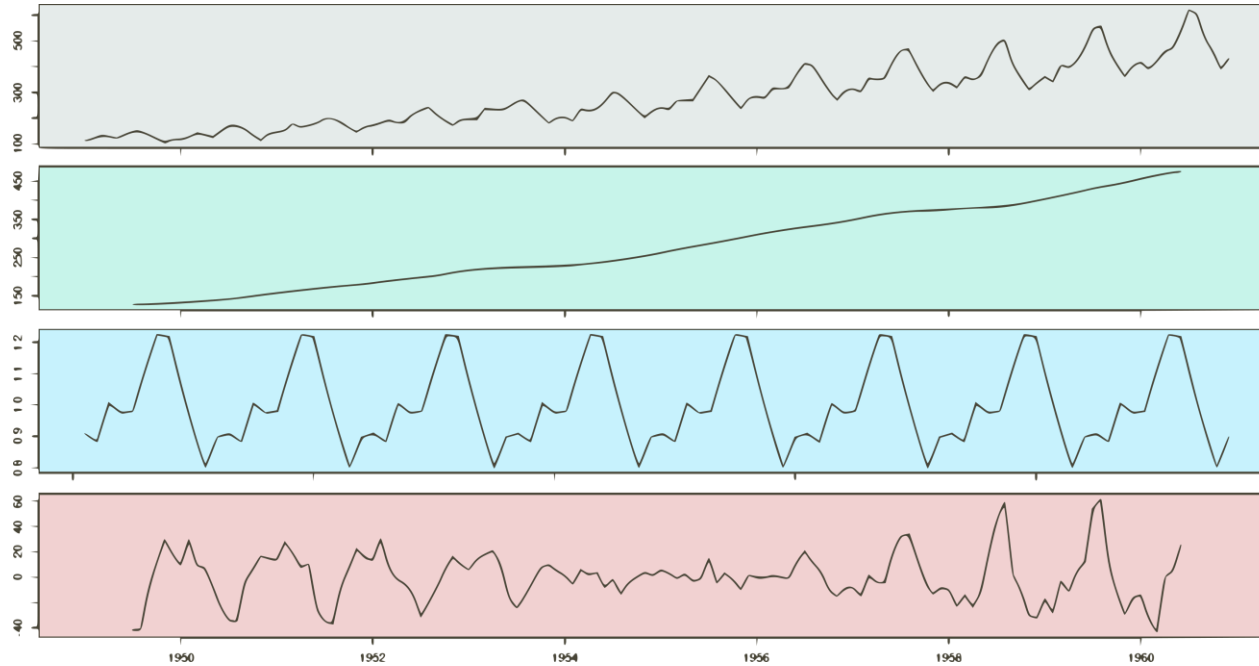
Construir una ruta óptima para un momento

VS

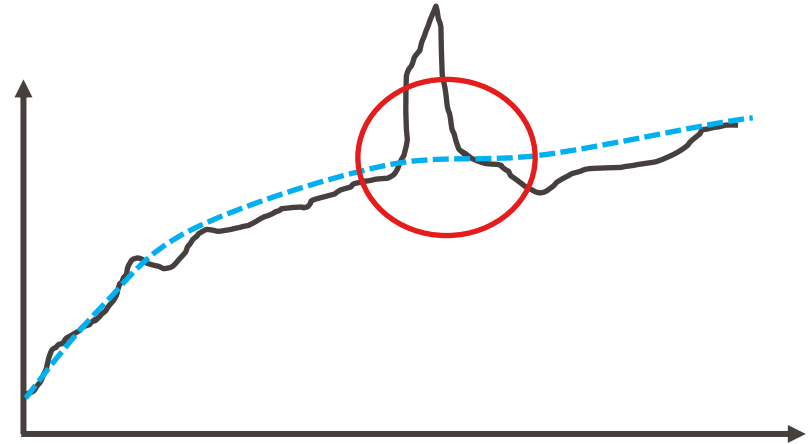
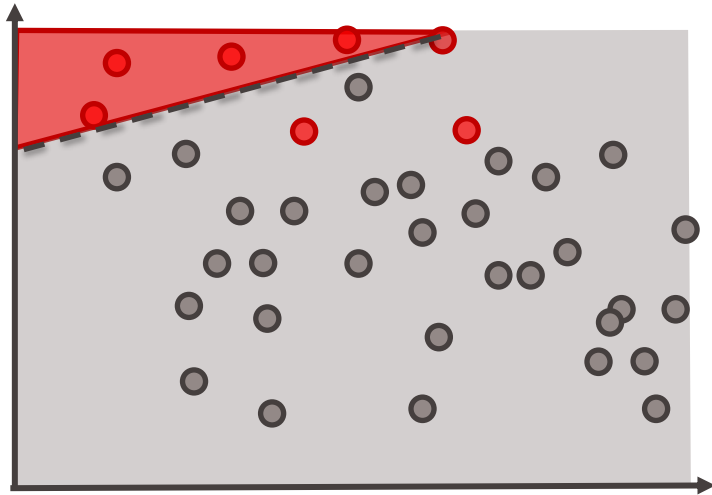
Construir rutas en pro medio *buenas* en todo momento



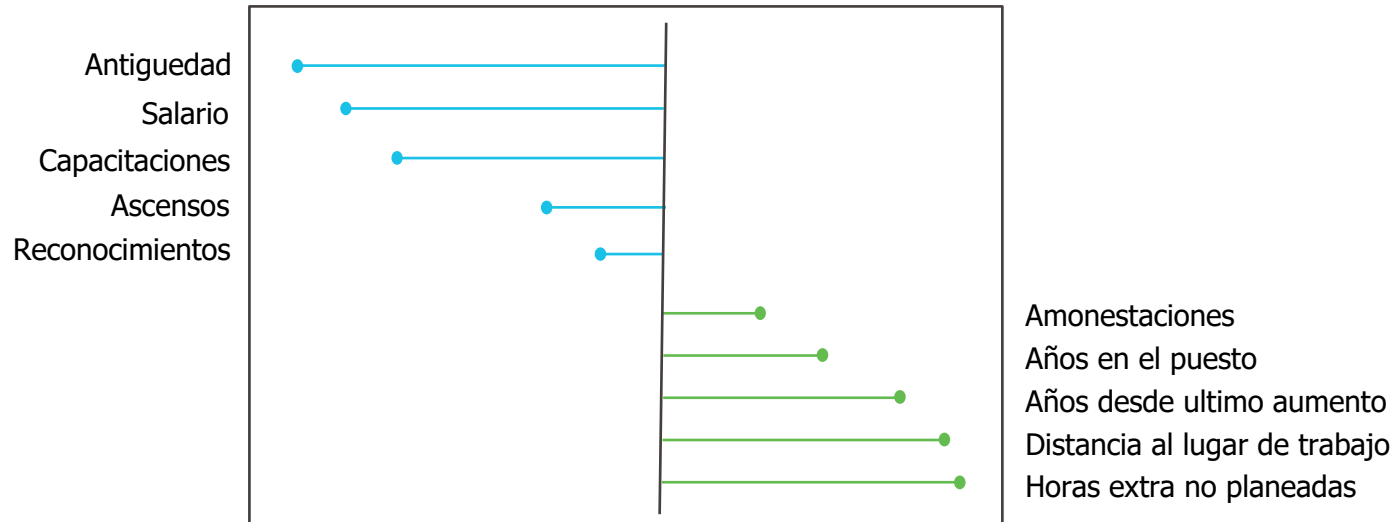
Modelación de demanda



Detección de anomalías

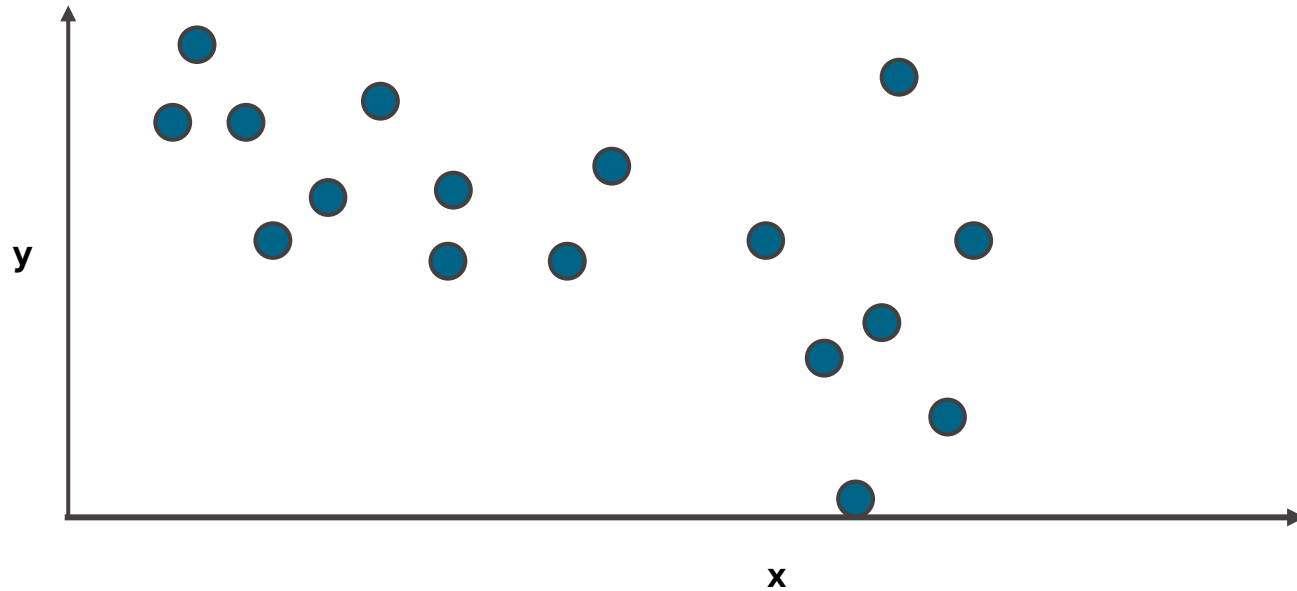


Rotación de personal

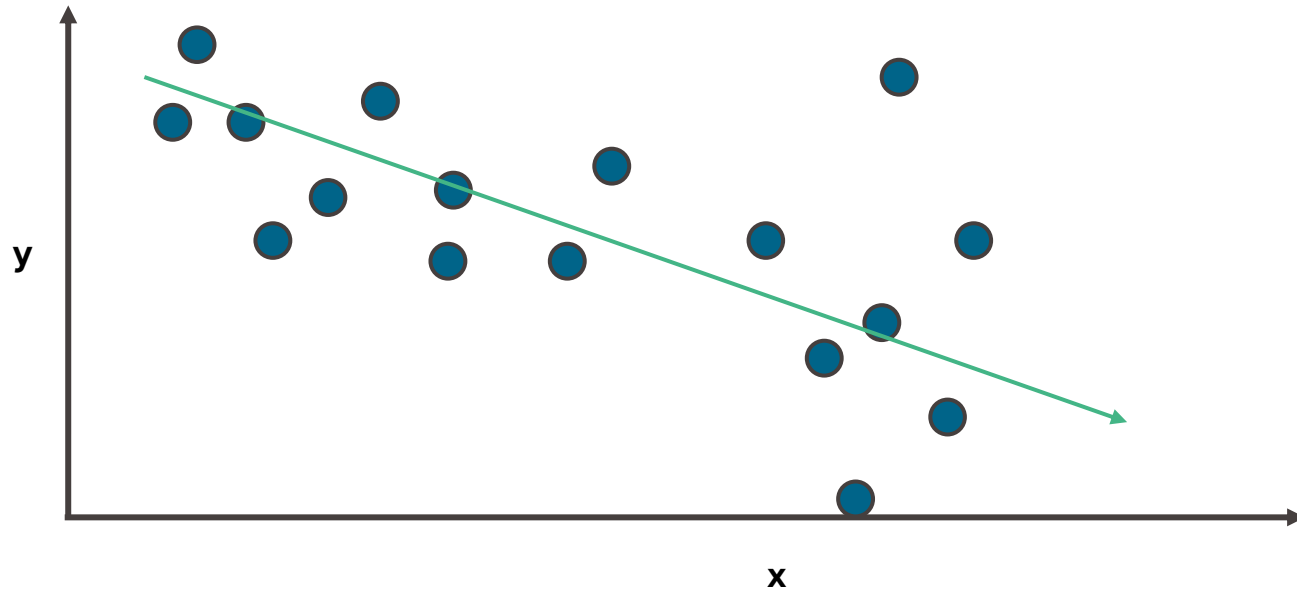


Modelo: Regresión Logística

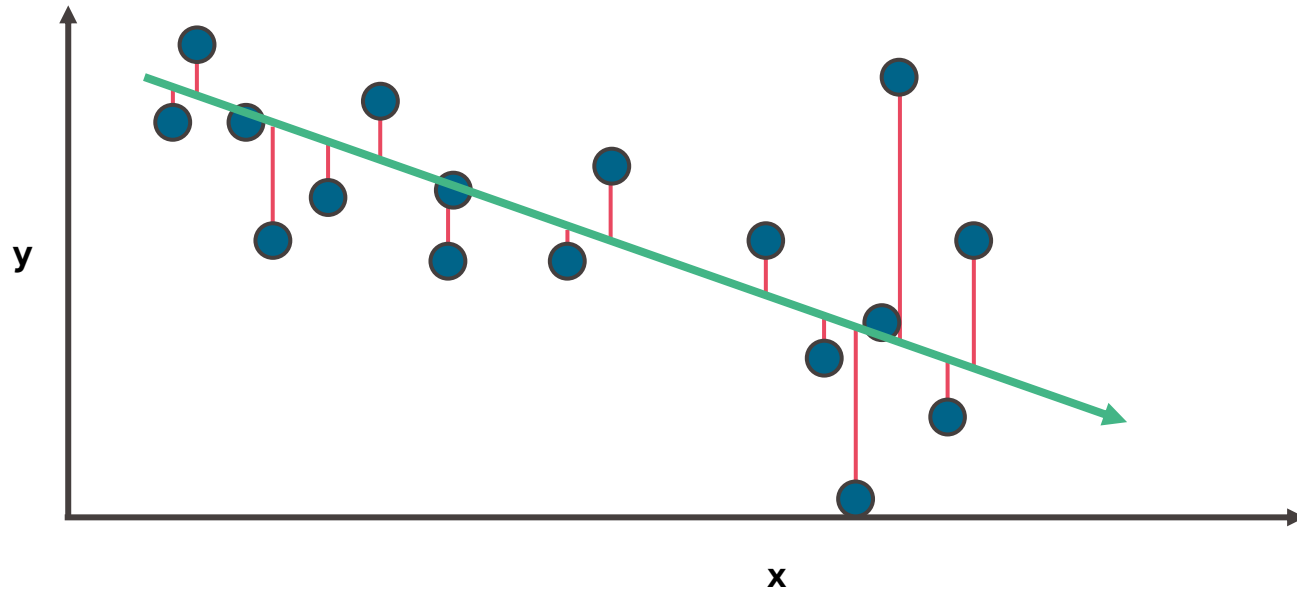
Regresión Lineal



Regresión Lineal

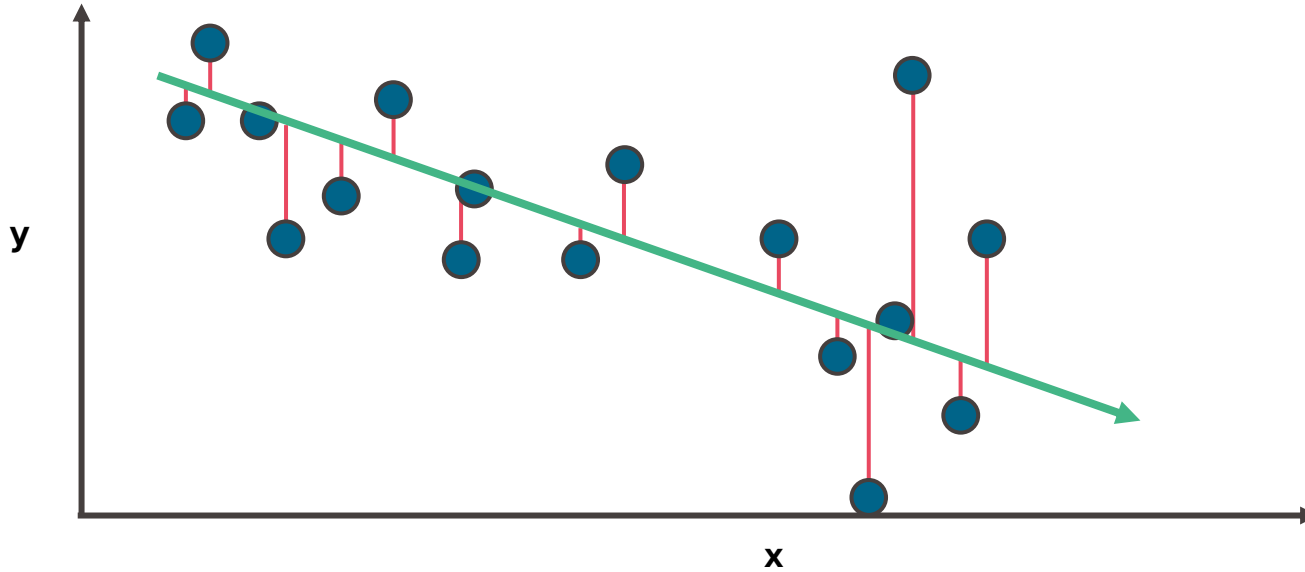


Regresión Lineal

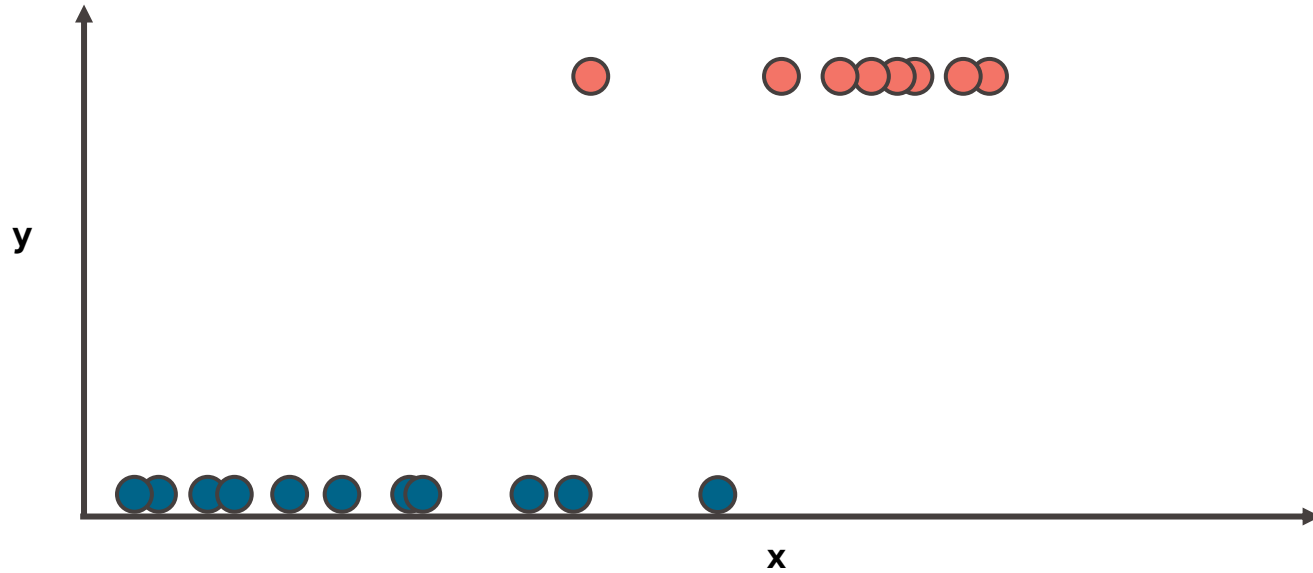


Regresión Lineal

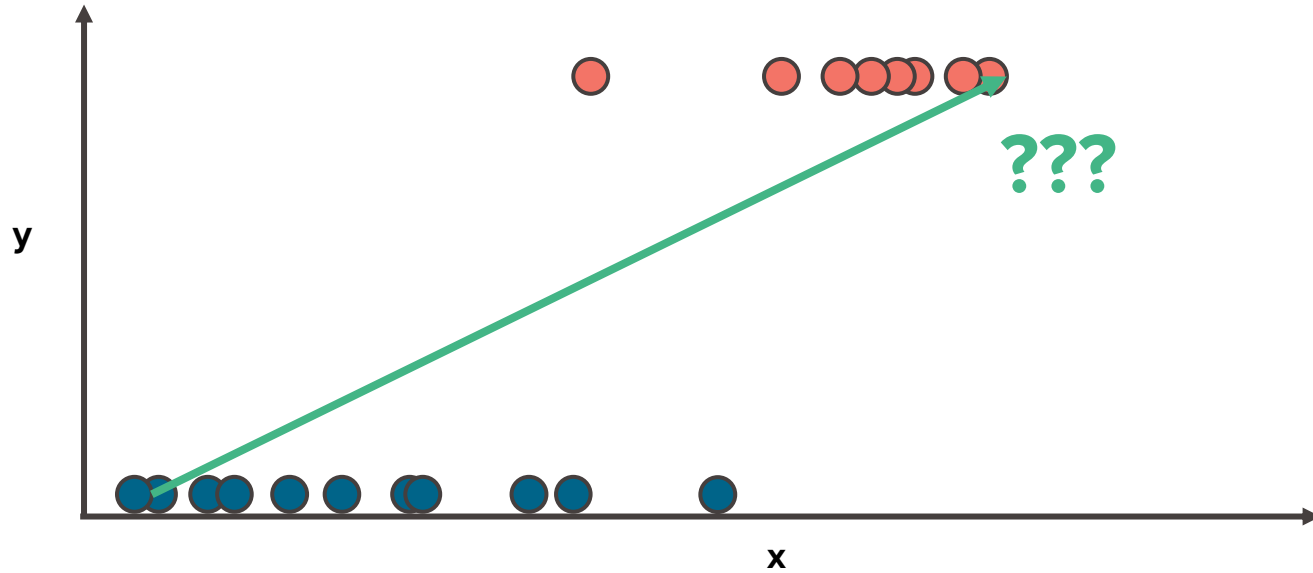
$$\min(\text{red square} + \text{red square} + \dots + \text{red square} + \text{red square})$$



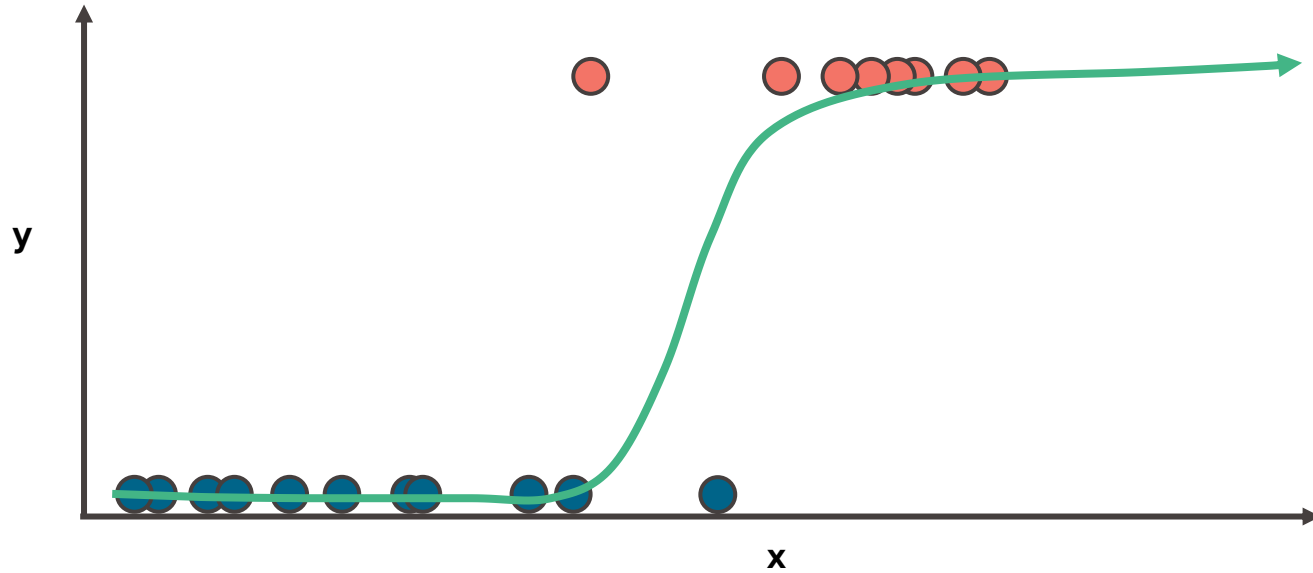
Regresión logística



Regresión logística



Regresión logística



Gracias!

Facebook

@gaplatam

Instagram

@gaplatam

Linkedin

Growth Acceleration Partners

Youtube

Growth Acceleration Partners LATAM

Demo

Aplicación: ejercicio.csv