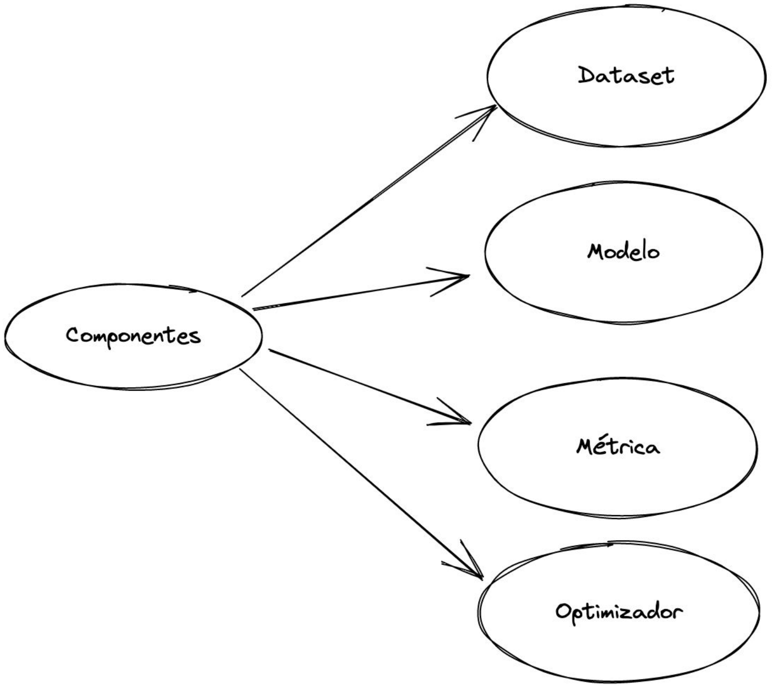
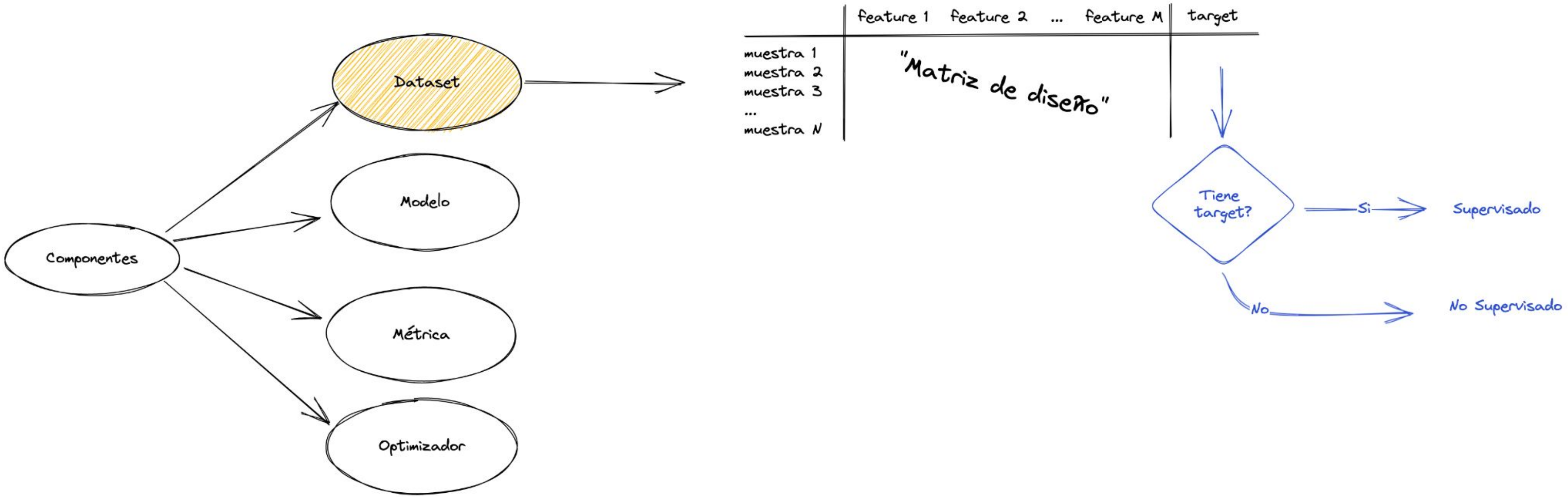
IAA-2023c1
Clase 2: ML End to End & Regresión Lineal

UNSAM
UNIVERSIDAD
NACIONAL DE
SAN MARTÍN

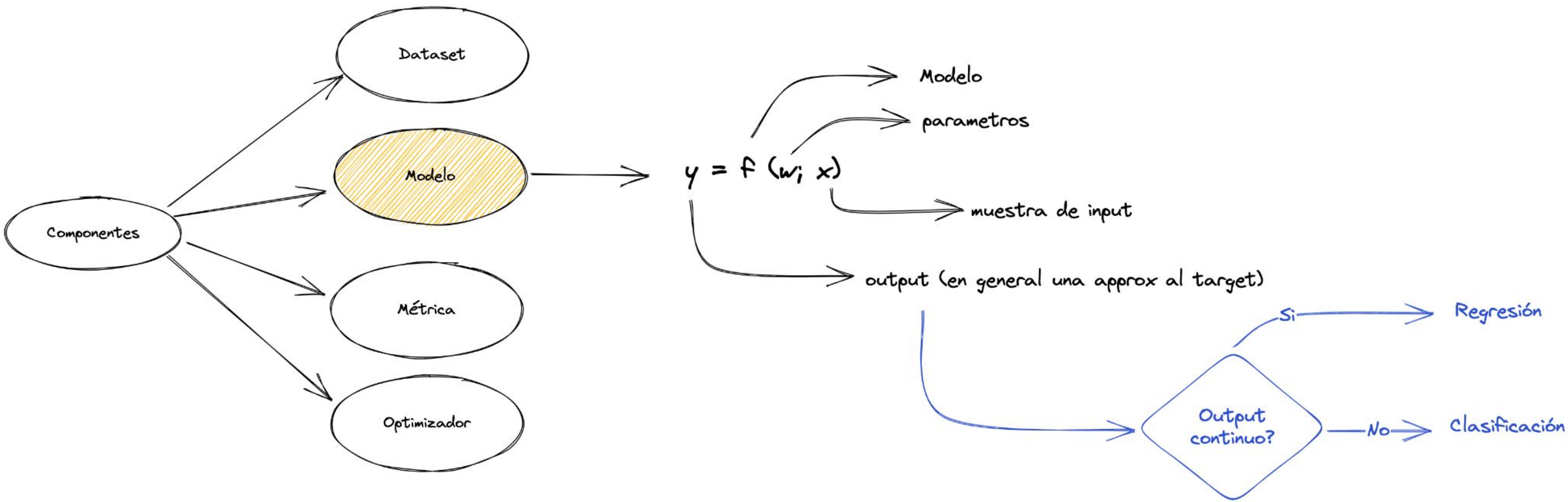
Repaso: Componentes



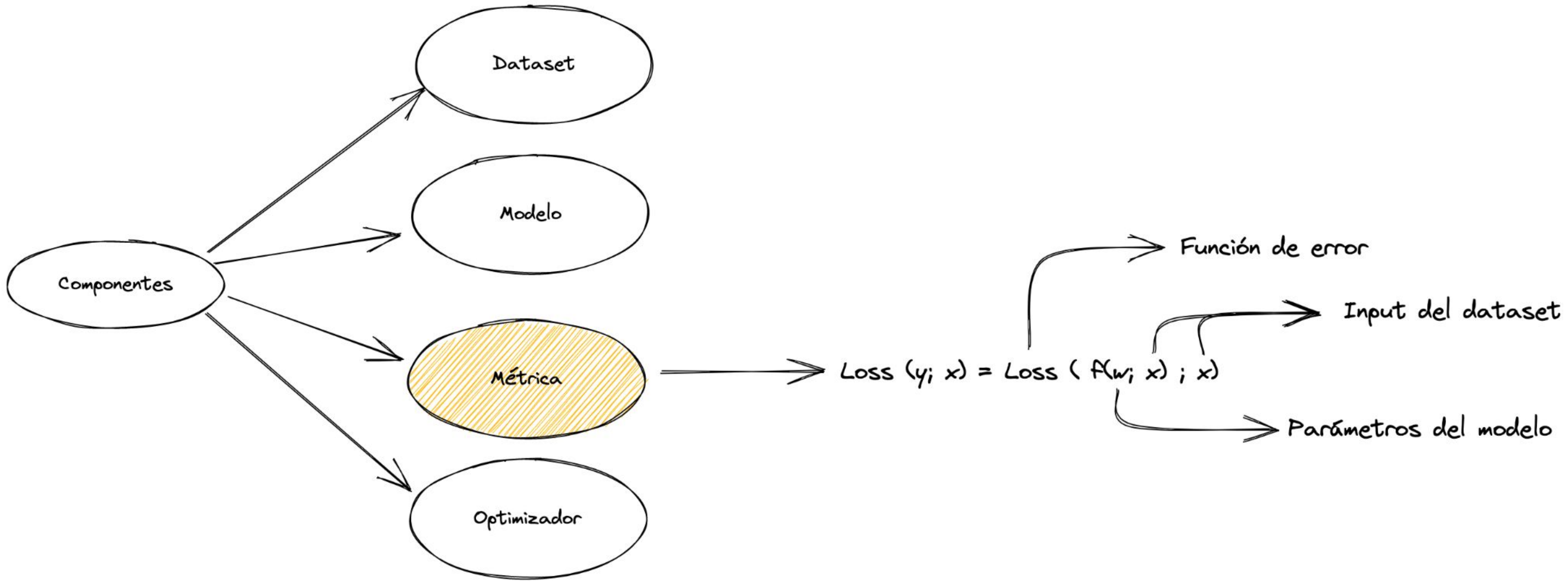
Repaso: Supervisado vs No Supervisado



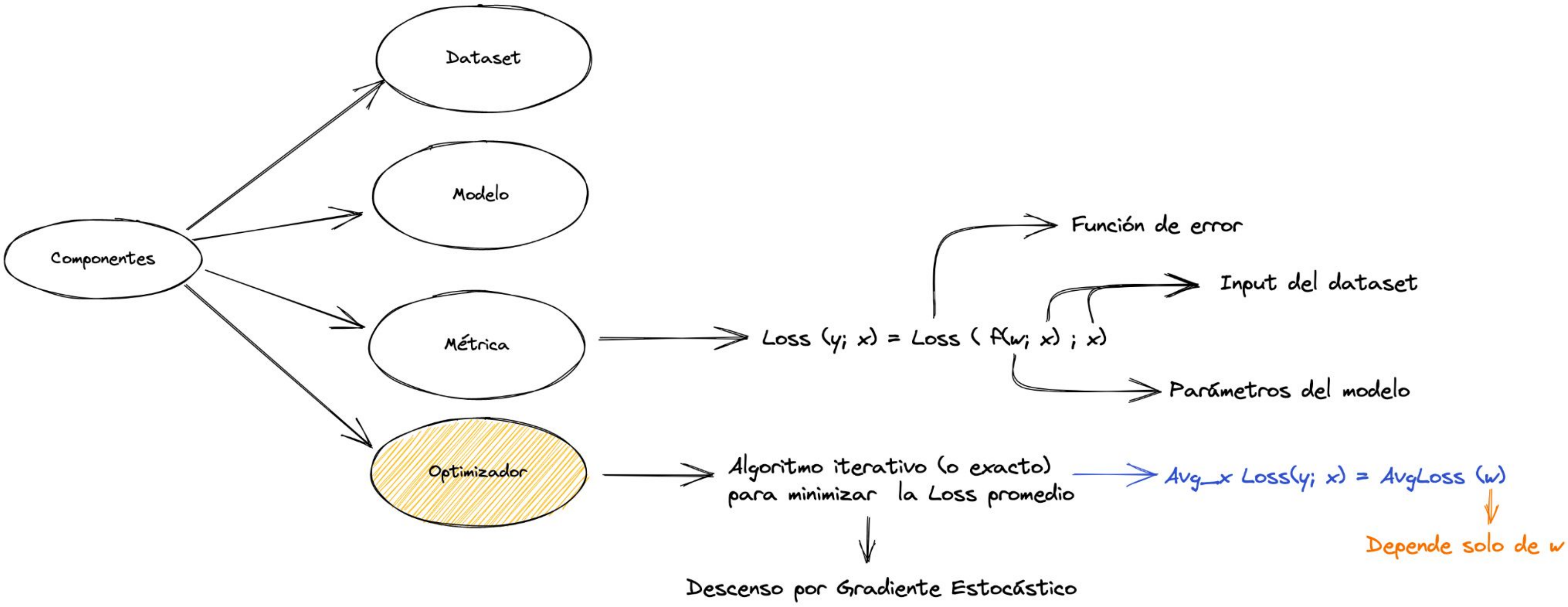
Repaso: Regresión vs Clasificación



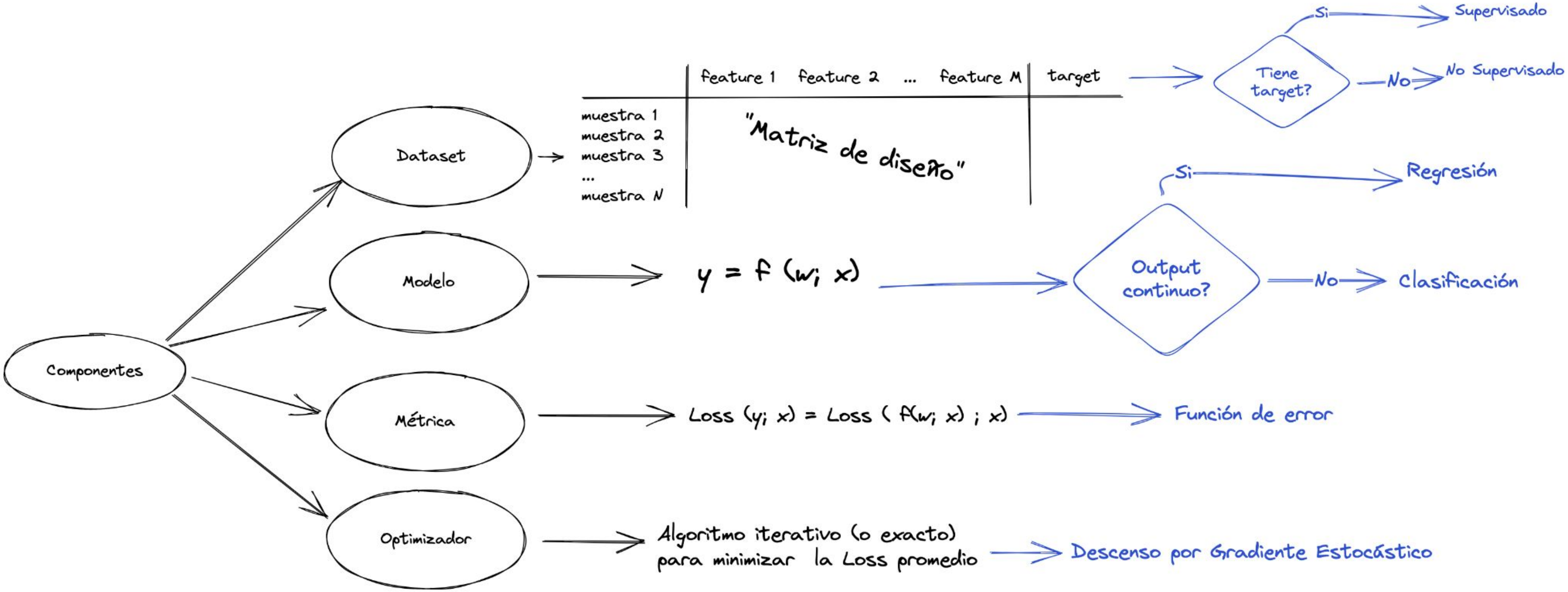
Repaso: Métrica y Función de Error / Pérdida



Repaso: Optimizador

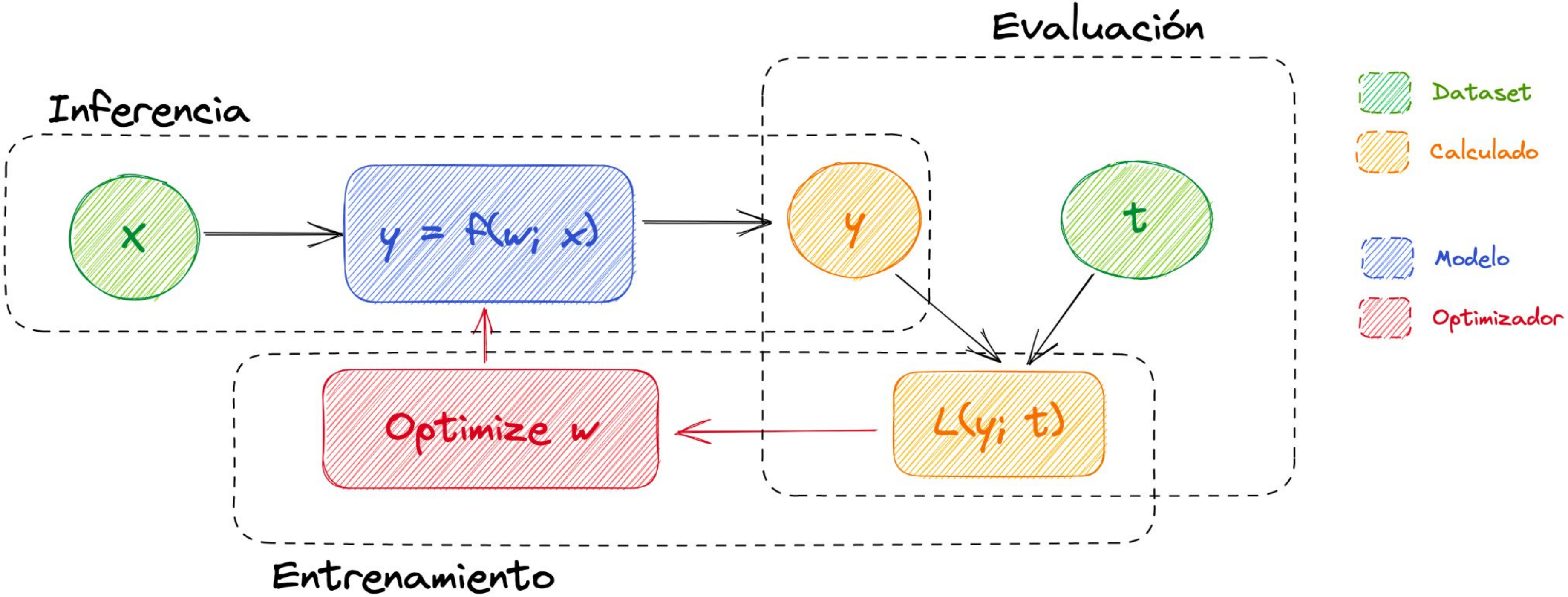


Repaso:

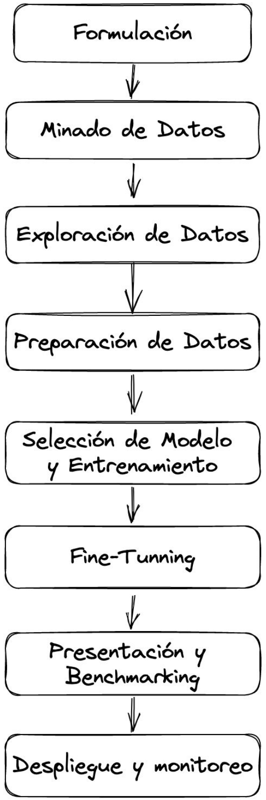


Componentes
Dataset
Modelo
Métrica
Optimizador
feature 1 feature 2 ... feature M
target
muestra 1
muestra 2
muestra 3
...
muestra N
"Matriz de diseño"
Tiene target?
Si
Supervisado
No
No Supervisado
y = f (w; x)
Output continuo?
Si
Regresión
No
Clasificación
Loss (y; x) = Loss (f(w; x) ; x)
Función de error
Algoritmo iterativo (o exacto)
para minimizar la Loss promedio
Descenso por Gradiente Estocástico

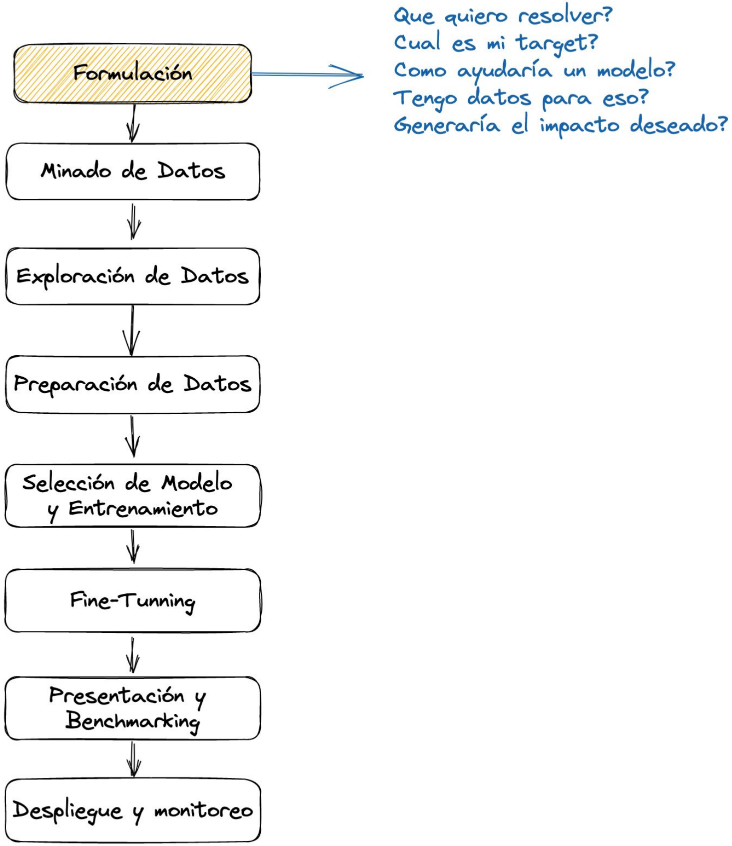
Repaso: Inferencia vs Entrenamiento



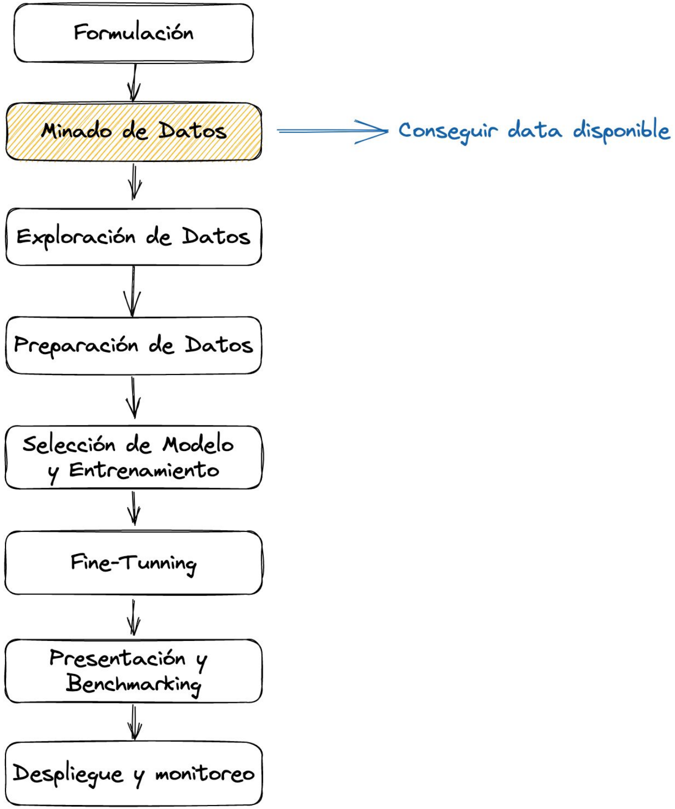
ML End to End


Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo

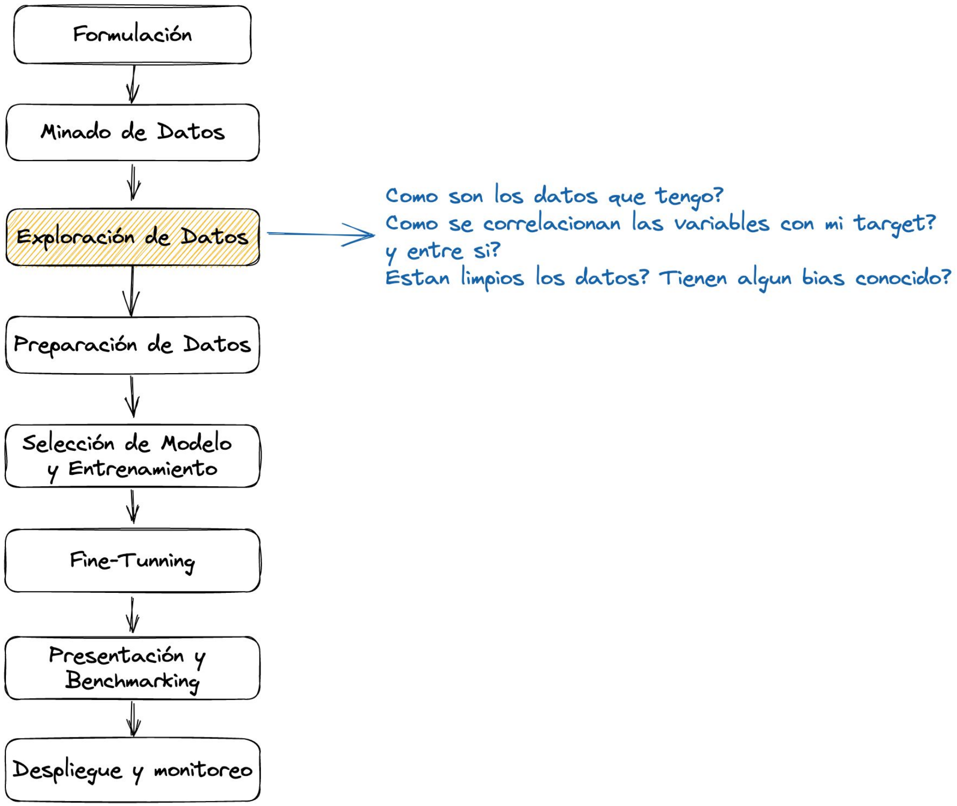
ML End to End


Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo
Que quiero resolver?
Cual es mi target?
Como ayudaría un modelo?
Tengo datos para eso?
Generaría el impacto deseado?

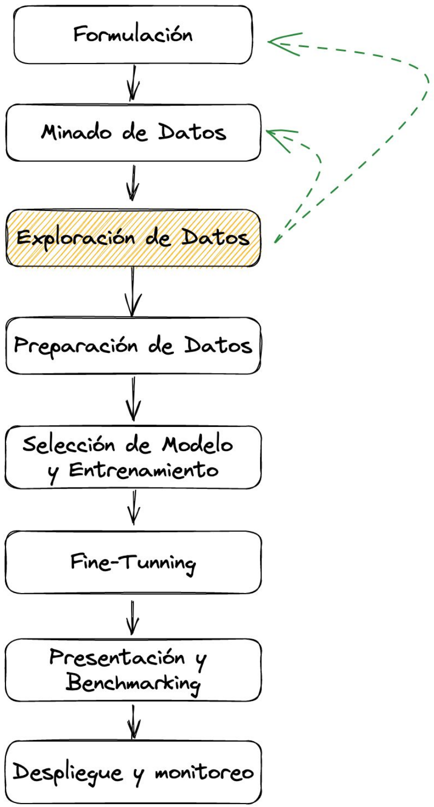
ML End to End


Formulación
Minado de Datos
Conseguir data disponible
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo

ML End to End

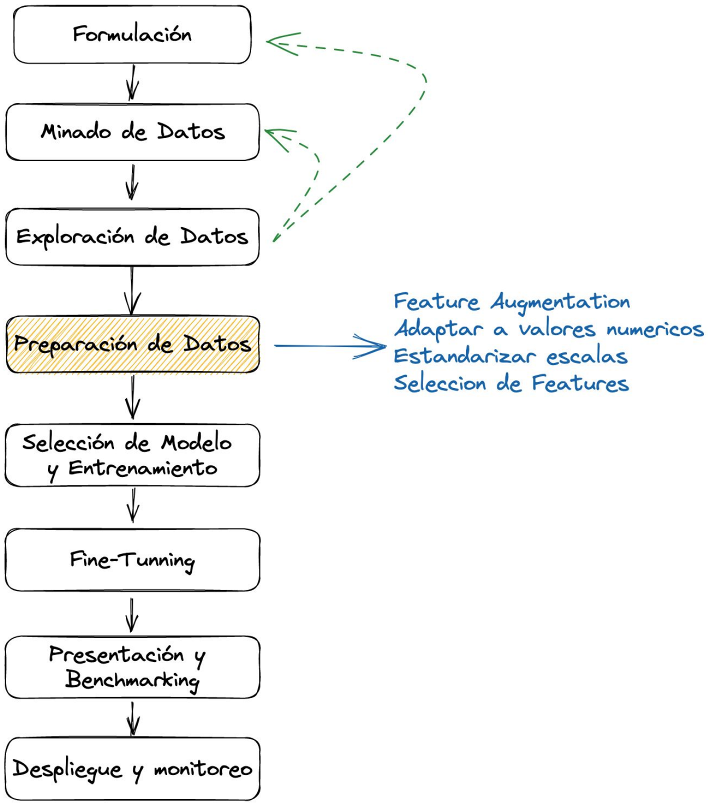

Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo
Como son los datos que tengo?
Como se correlacionan las variables con mi target?
y entre si?
Estan limpios los datos? Tienen algun bias conocido?

ML End to End

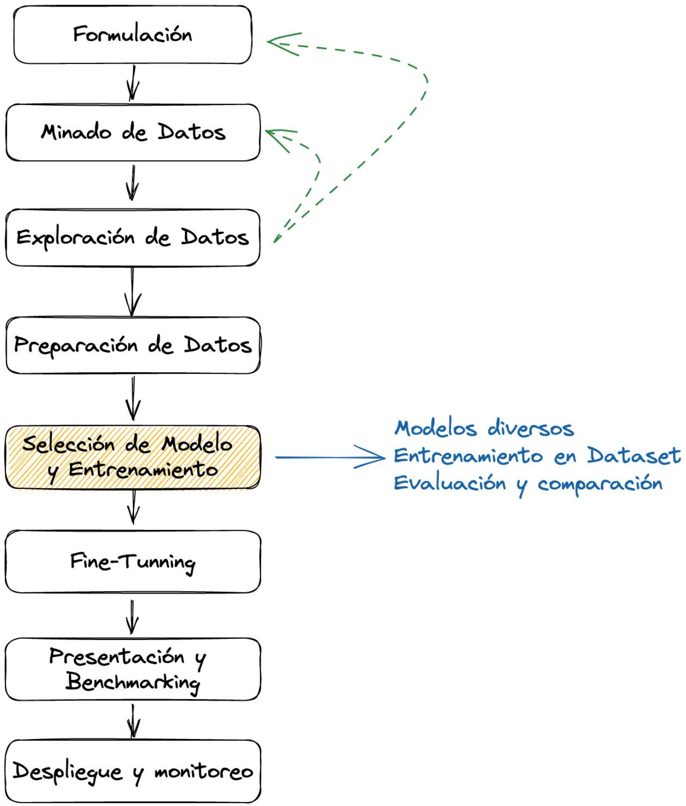


Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo

ML End to End

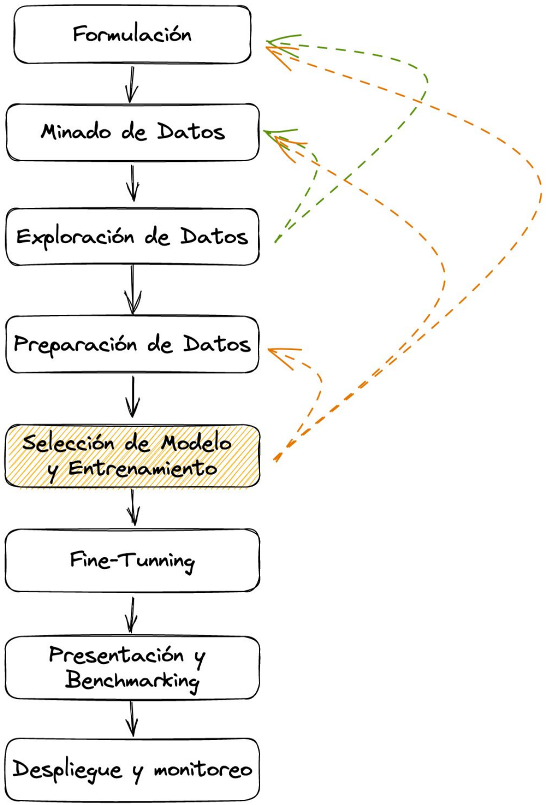

Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo
Feature Augmentation
Adaptar a valores numericos
Estandarizar escalas
Seleccion de Features

ML End to End

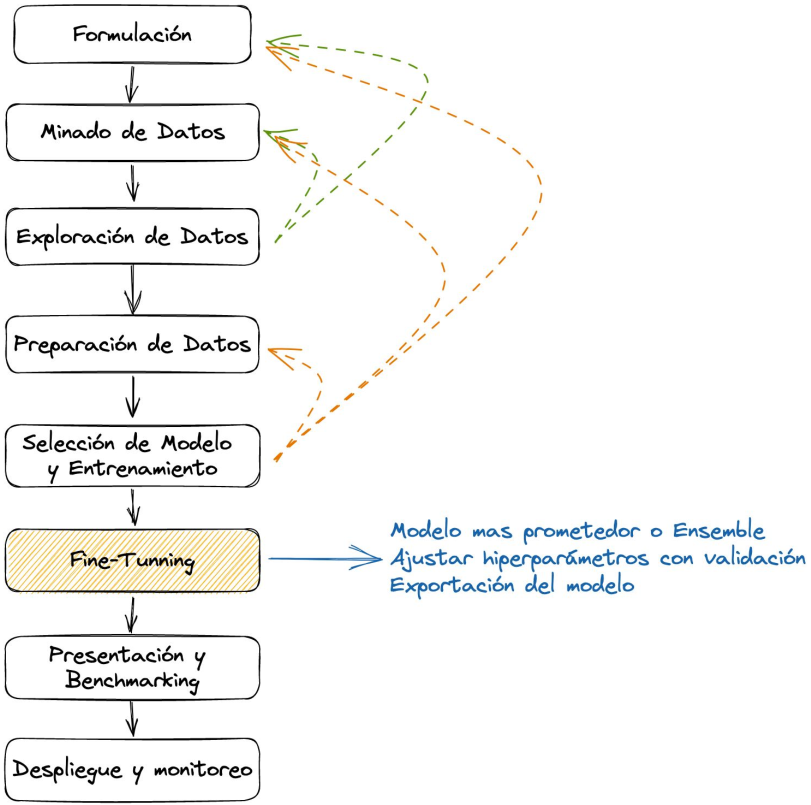


Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Modelos diversos
Entrenamiento en Dataset
Evaluación y comparación
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo

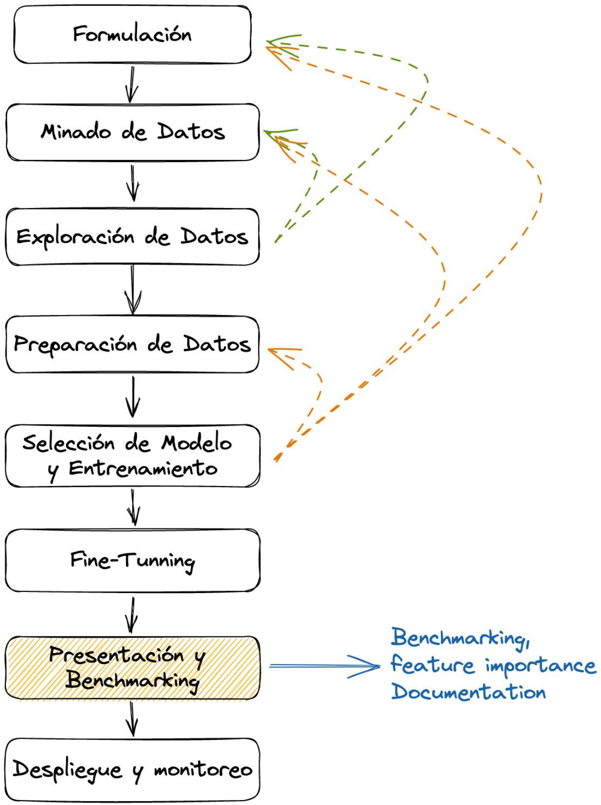
ML End to End


Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo

ML End to End

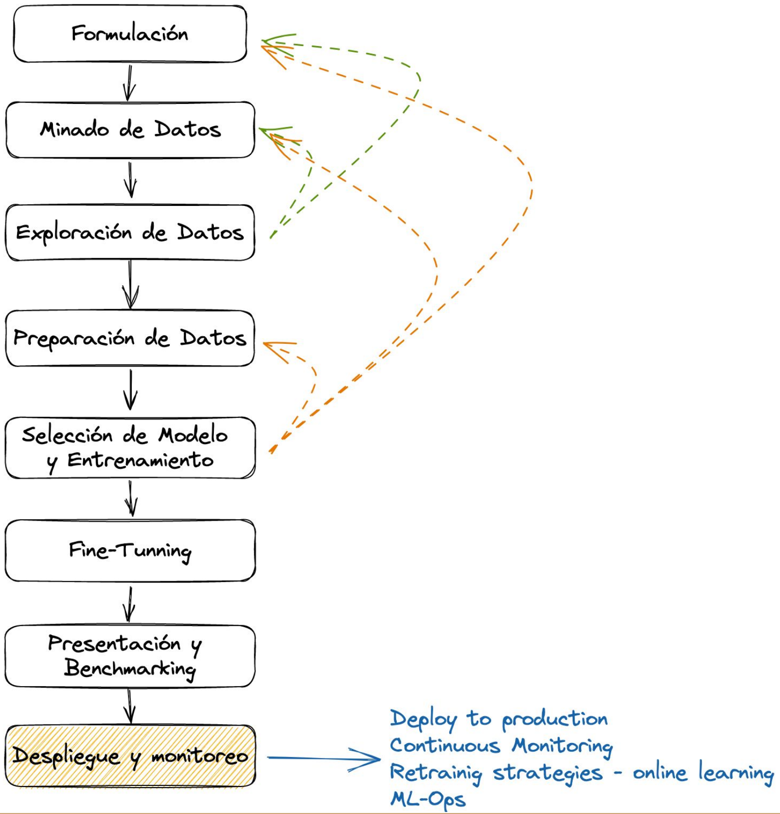

Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo
Modelo mas prometedor o Ensemble
Ajustar hiperparámetros con validación
Exportación del modelo

ML End to End



Formulación
Minado de Datos
Exploración de Datos
Preparación de Datos
Selección de Modelo y Entrenamiento
Fine-Tunning
Presentación y Benchmarking
Despliegue y monitoreo
Benchmarking,
feature importance
Documentation

ML End to End



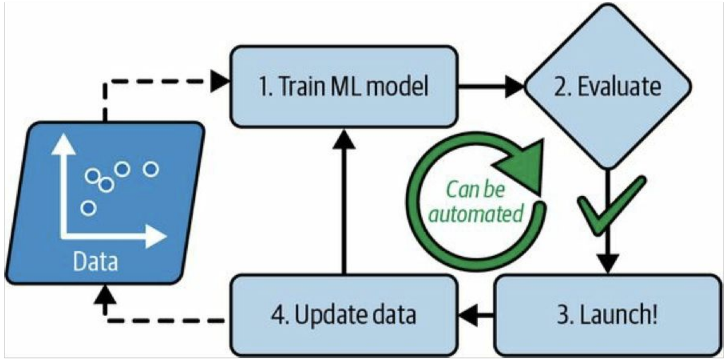
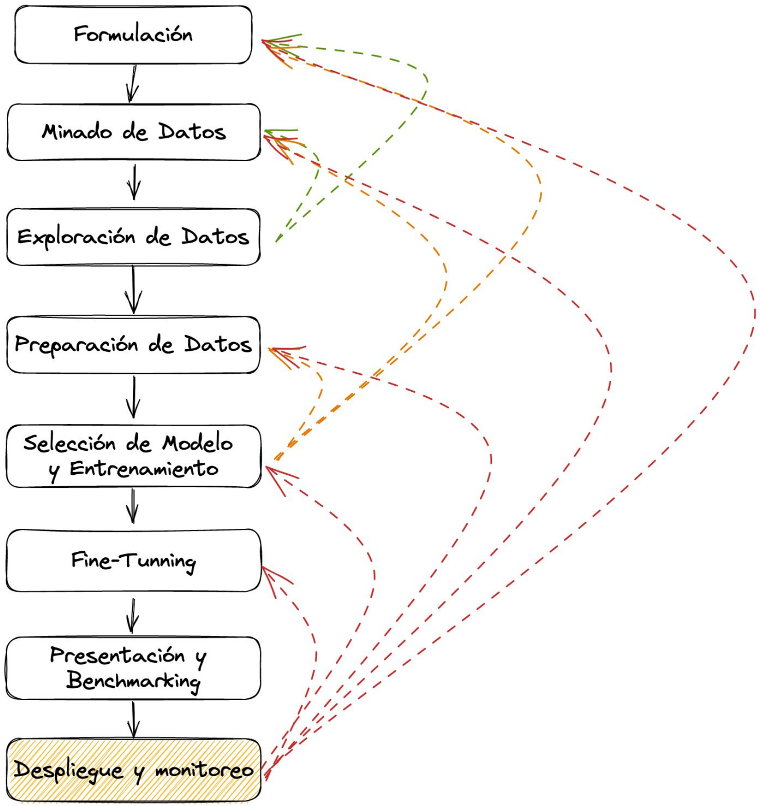


Figure 1-3. Automatically adapting to change

ML End to End



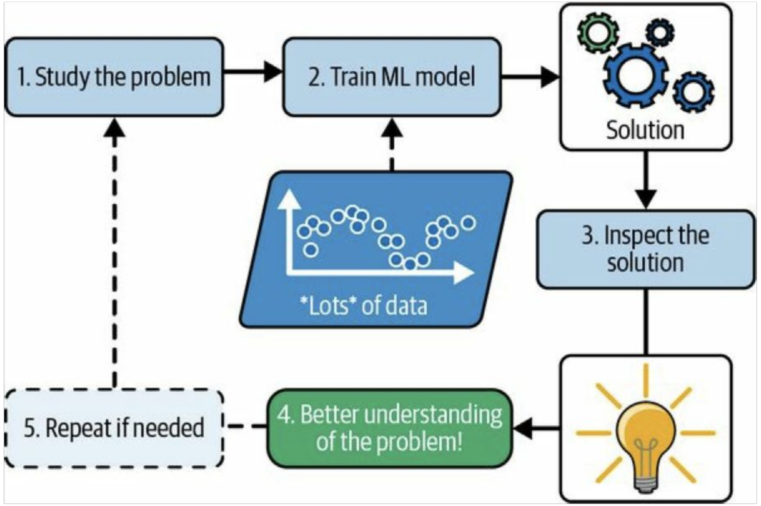
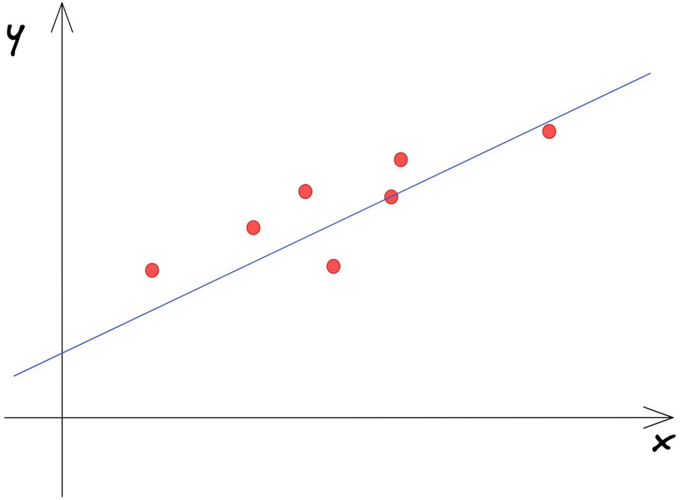


Figure 1-4. Machine learning can help humans learn

Regresión: Modelo Lineal

- Ecuación de una recta:

$$y = ax + b$$



Regresión: Modelo Lineal

- Ecuación de una recta:

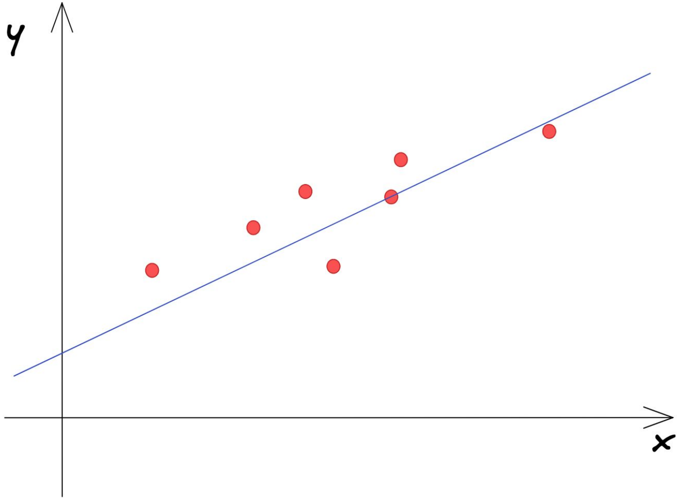
dep. var
a.k.a input

$$y = ax + b$$

ind. var
a.k.a. output

Pendiente
a.k.a. peso

Ordenada al origen
a.k.a. bias



Regresión: Modelo Lineal

- Ecuación de una recta:

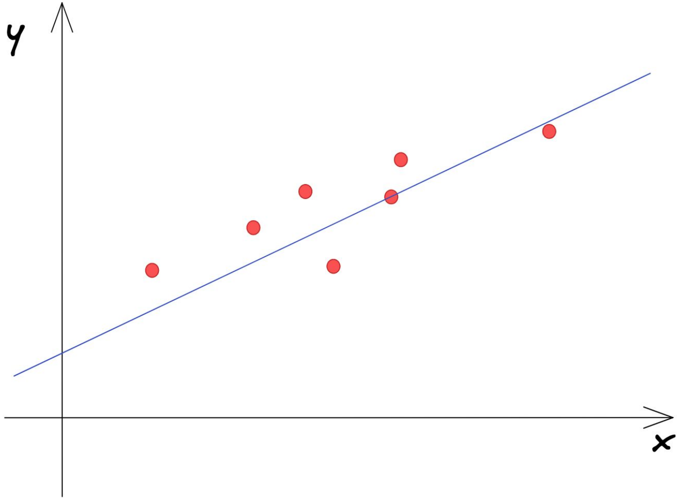
dep. var
a.k.a input

$$y = ax + b$$

ind. var
a.k.a. output

Pendiente
a.k.a. peso

Ordenada al origen
a.k.a. bias



$$y = w_1 x + w_0$$

pesos

Regresión: Modelo Lineal

- Una sola variable: Un solo feature

$$y = w_1 x + w_0$$

- Multivariable: Muchas features

$$y = w_0 + w_1 x_1 + \cdots + w_N x_N$$

$$= \begin{pmatrix} w_0 \\ w_1 \\ \vdots \\ w_N \end{pmatrix} \begin{pmatrix} 1 & x_1 & \dots & x_N \end{pmatrix}$$

Regresión: Modelo Lineal

- Multivariable:
  Muchas features
  Un solo sample

$$y = \begin{pmatrix} w_0 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} \begin{pmatrix} 1 & x_1 & x_2 & \dots & x_N \end{pmatrix} = \vec{w} \cdot \vec{x}$$

- Multivariable:
  Muchas features
  Muchos Samples

$$\vec{y} = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(M)} \end{pmatrix} = \begin{pmatrix} w_0 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \dots & x_N^{(1)} \\ 1 & x_1^{(2)} & x_2^{(2)} & \dots & x_N^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(M)} & x_2^{(M)} & \dots & x_N^{(M)} \end{pmatrix}$$

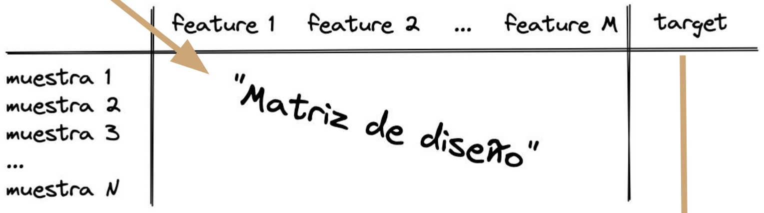
$$= \vec{w} \cdot \bar{X}$$

Regresión: Modelo Lineal

- Modelo lineal $\vec{y} = \vec{w} \cdot \bar{X}$

Output / Predicción

Parámetros



target: $\vec{t}$

- ¿Qué más hace falta?

Regresión Lineal: Función de Pérdida/Error

- Error cuadrático

$$
\begin{aligned}
e(y^{(i)}); t^{(i)}) &= (y^{(i)} - t^{(i)})^2 \\
&= (y^{(i)}(\vec{w}; \vec{x}^{(i)}) - t^{(i)})^2
\end{aligned}
$$

Regresión Lineal: Función de Pérdida/Error

- Error cuadrático

$$
\begin{aligned}
e(y^{(i)}); t^{(i)}) &= (y^{(i)} - t^{(i)})^2 \\
&= (y^{(i)}(\vec{w}; \vec{x}^{(i)}) - t^{(i)})^2
\end{aligned}
$$

- Error cuadrático medio (MSE)

$$
\begin{aligned}
MSE(\vec{y}; \vec{t}) &= \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)})^2 \\
&= \frac{1}{N} \left| \vec{y} - \vec{t} \right|^2 \\
&= MSE(\bar{X}, \vec{t}; \vec{w})
\end{aligned}
$$

Regresión Lineal: Optimizador

- Minimizar el error:
  Queremos encontrar los pesos que minimizan el la hiper-paraboloide

$$MSE(\bar{X}, \vec{t}; \vec{w}) = \frac{1}{N} \left| \vec{w} \cdot \bar{X} - \vec{t} \right|^2$$

- Gradient Descent:
  Nos podemos mover en la dirección de máxima variación del gradiente

$$\nabla_{\vec{w}} MSE = \frac{2}{N} (\vec{w} \cdot \bar{X} - \vec{t}) \cdot \bar{X}^t$$

- Solver: En este caso hay una solución algebraica exacta al gradiente nulo (mínimo).

Regresión Lineal

- Modelo

$$\vec{y} = \vec{w} \cdot \bar{X}$$

- Función de Pérdida MSE

$$MSE(\bar{X}, \vec{t}; \vec{w}) = \frac{1}{N} \left| \vec{w} \cdot \bar{X} - \vec{t} \right|^2$$

- Optimizador:
  - Solver algebraico (exacto)
  - ó Descenso por Gradiente (iterativo)

$$\nabla_{\vec{w}} MSE = \frac{2}{N} (\vec{w} \cdot \bar{X} - \vec{t}) \cdot \bar{X}^t$$