



Fundamentos de Data Science con Python

Sesión 3: Introducción a Machine Learning

Dr. Julio Lopez-Nunez

Diciembre, 2025



Objetivos de la sesión.



Comprender qué es
Machine Learning y sus
tipos.




Diferenciar entre
regresión y clasificación.





Entrenar modelos
simples con scikit-learn.



Evaluar la calidad de los
modelos (métricas
básicas).



¿Qué es Machine Learning?



Definición: sistemas que aprenden de los datos para hacer predicciones.



Se diferencia de la programación tradicional: no hay reglas explícitas → el modelo aprende patrones.



Ejemplos:



Educación: predecir riesgo de deserción.



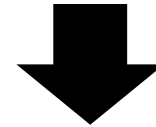
Negocios: segmentar clientes.



Salud: diagnóstico a partir de imágenes.

Tipos de Machine Learning.

Tipos ML



Supervisado

Regresión (predice números).

Clasificación (predice categorías).

No supervisado

Clustering (agrupa datos sin etiquetas).

Reducción de dimensionalidad.

Refuerzo (menor detalle)

Aprendizaje por recompensas.

Ciclo de modelado en ML.

se repite hasta lograr un modelo adecuado



Regresión lineal.

Objetivo: predecir un valor numérico.

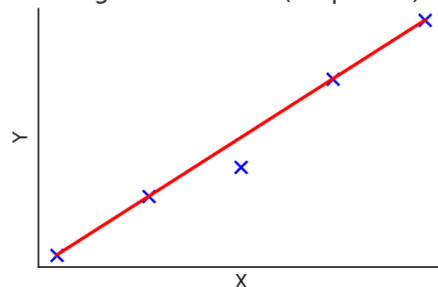
Ejemplo: nota final en función de asistencia.

Fórmula: $y = b_0 + b_1x_1 + b_2x_2 + \dots$

Librería: LinearRegression de scikit-learn.

Métrica: RMSE (Root Mean Squared Error).

Regresión Lineal (esquema)



Clasificación.

Objetivo: predecir categorías.

Ejemplo: predecir si un estudiante está Aprobado/Reprobado.

Algoritmos introductorios:

Regresión logística, Árbol de decisión.

Métricas:

Accuracy (exactitud), Matriz de confusión, F1-score y Precisión/Recall, entre otras.



Underfitting vs Overfitting.



Underfitting:

El modelo es demasiado simple, no logra capturar los patrones de los datos.

Ejemplo: una recta intentando ajustar datos con forma curva.



Overfitting:

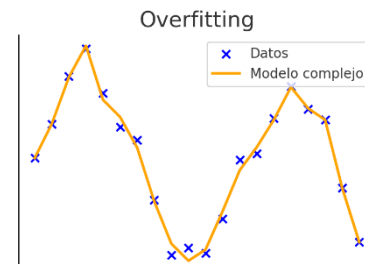
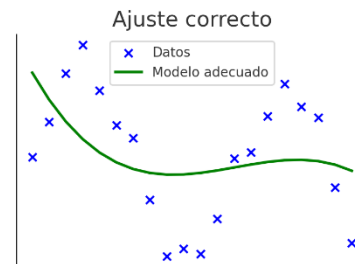
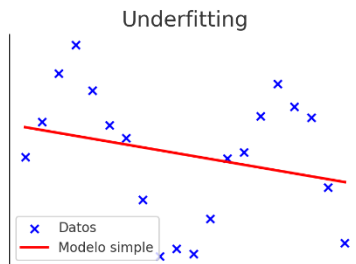
El modelo es demasiado complejo, se ajusta en exceso a los datos de entrenamiento, pero falla al generalizar.

Ejemplo: curva que sigue punto por punto al dataset.



Ajuste adecuado (just right):

El modelo encuentra el balance, explica los patrones sin memorizar el ruido.



Seleccionar el Mejor Modelo/Algoritmo.



No existe una única métrica universal:
depende del tipo de problema, y algoritmo
utilizado.



Algunas Métricas:

- Regresión → RMSE, MAE, R^2 .
- Clasificación → Accuracy, Precision, Recall, F1, AUC-ROC.
- Clustering → Silhouette, Davies-Bouldin, Calinski-Harabasz.

□ Métricas de Evaluación en ML (resumen)



Problema	Métrica	Qué mide / Rango
Regresión	RMSE	Error promedio ($0 \rightarrow \infty$)
	R^2	Varianza explicada ($-\infty \rightarrow 1$)
Clasificación	Accuracy	% aciertos ($0 \rightarrow 1$)
	F1-Score	Balance Precisión/Recall
Clustering	Silhouette	Cohesión/separación ($-1 \rightarrow 1$)



Actividad práctica.



Importar dataset (trabajamos con el CSV `datos_estudiantes_practicas.csv`).

Explorar filas y columnas.

Seleccionar Variables (Objetivo y Predictoras)

Generar Modelo y Evaluar



Discusión grupal.



¿Qué es realmente Machine Learning?

¿Una técnica? ¿Una herramienta? ¿O una forma distinta de pensar los datos?



¿Dónde termina la automatización y comienza la experiencia del analista?