

guayverd

Fundamentos IA

Machine Learning

Clase 10

En colaboración con
IBM SkillsBuild





¡Bienvenidos!

¿Nos presentamos?

- ¿Qué recuerdan de la clase anterior?
- ¿Qué esperan aprender?
- ¿Tienen alguna pregunta?

Contenidos

Por temas

09

- Demo

10

- Fundamentos

11

- Scikit-learn
- Demo asincrónica

12

- Introducción al entorno
- Modelado y relaciones

Objetivos de la clase



- Machine Learning
- Tipos de aprendizajes
- Algoritmos básicos
- Métricas de evaluación

Machine Learning

Fundamentos

Plataforma Skill Build: Machine Learning



Plan de formación

Python for Machine Learning: Unlocking the Power of Artificial Intelligence

🕒 Aproximadamente 23 horas 🎓 281 ★★★★★ 18



eLearning

Modelos de IBM Granite para el desarrollo de softwar

1 hora 30 minutos 🎓 8.825 ★★★★★ 623



eLearning

Clasificación de datos con IBM Granit

1 hora 30 minutos 🎓 10.885 ★★★★★ 597



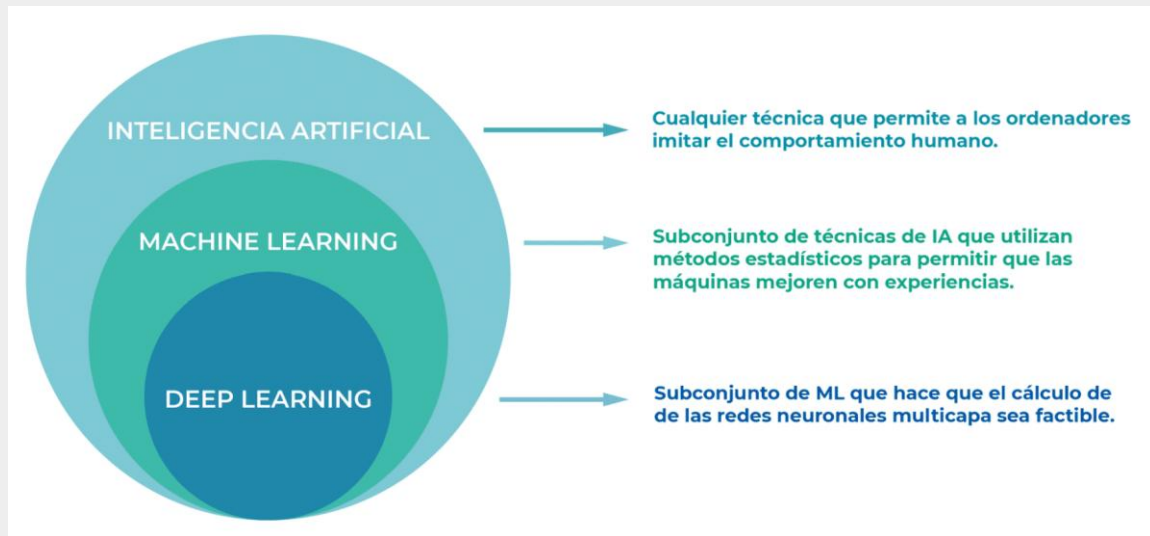
¿Qué es Inteligencia Artificial?

Es una rama de las ciencias de la Computación que diseña y crea entidades con la capacidad de percibir datos de su entorno, analizarlos, asimilarlos y utilizarlos para conseguir un objetivo; de forma semejante a las capacidades humanas de cognición, y razonamiento.



¿Qué es Machine Learning?

ML se inserta en una disciplina más general, macro, conocida como **Inteligencia Artificial**.



¿Qué es Machine Learning?

Algoritmos que **aprenden patrones de datos para hacer predicciones**

- Sistema que mejora automáticamente con la experiencia
- Encuentra patrones sin programación explícita
- Hace predicciones sobre datos nuevos
- Aplicación práctica de Inteligencia Artificial



¿Cómo aprenden las máquinas?

¿En qué se diferencia del aprendizaje humano?

Algoritmo en ML

¿cómo hacemos para que las computadoras aprendan de los datos?

- Debemos comenzar hablando de modelos matemáticos y cómo estos se relacionan con ML.
- Una primera aproximación a un modelado podría armarse con las observaciones y los patrones que se desprenden de eventos pasados, ya que podríamos aseverar que en el futuro podrían repetirse esos componentes y eventos

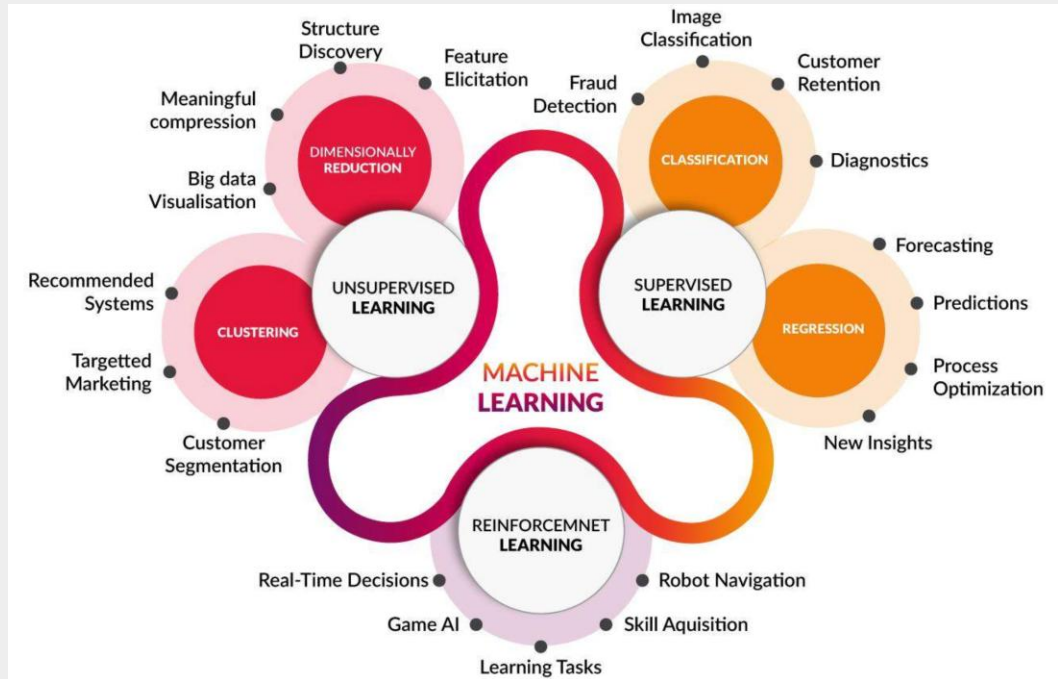
Conjunto de instrucciones que permiten:

- Aprender patrones a partir de datos
- Ajustar relaciones entre variables
- Generar predicciones o clasificaciones

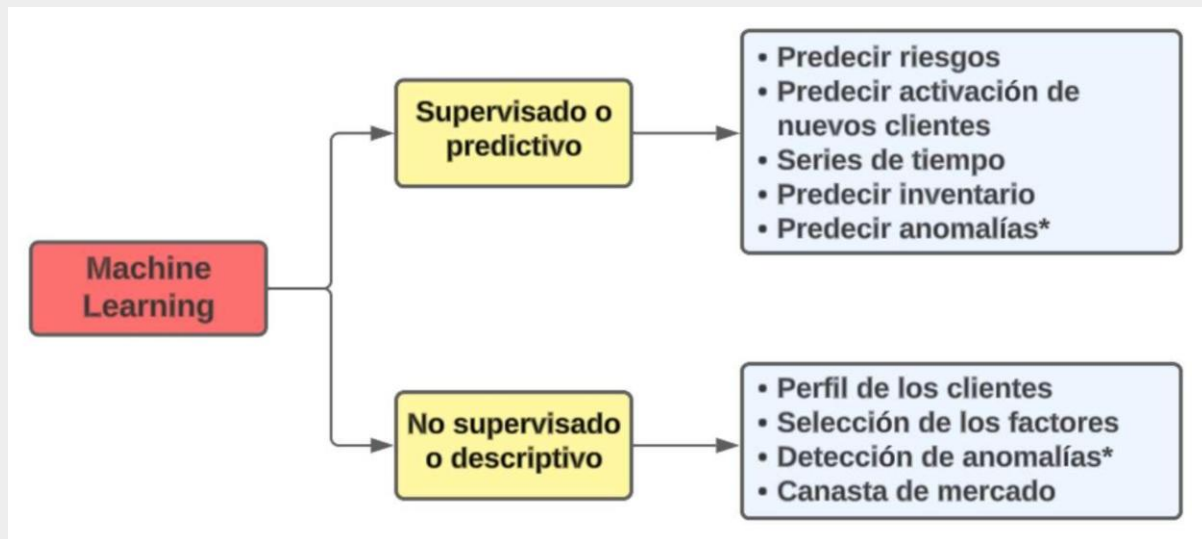
Es el procedimiento que guía el aprendizaje del modelo a partir de los datos

Tipos de ML

- Supervised
- Unsupervised
- Reinforcement



Tipos de aprendizaje



Tipos de aprendizaje

Supervisado

- Usa datos etiquetados (entrada–salida conocida)
- Aprende de ejemplos con respuestas correctas
- **Ejemplos:** clasificación de spam, predicción de precios

No supervisado

- Trabaja con datos sin etiquetas
- Descubre patrones o estructuras ocultas
- **Ejemplos:** agrupamiento de clientes, detección de anomalías

Algoritmos básicos

Supervisado

- **Regresión lineal:** predice valores numéricos
- **Clasificación binaria:** distingue entre dos categorías

No supervisado

- **Clustering:** agrupa datos según similitudes

Aprendizaje supervisado

Buscamos un modelo f que permita determinar la salida y a partir de la entrada X .

En esta función:

- X son los atributos (generalmente se denota con mayúscula porque incluye más de una variable)
- y es la etiqueta

El aprendizaje supervisado permite modelar la relación entre las características medidas de los datos y alguna etiqueta asociada con ellos.

Es decir, podremos predecir y para nuevos datos X de los cuales no conozcamos la salida.

Regresión (supervisado)

La variable de salida es un valor numérico. Ejemplo: precio, cantidad, temperatura.

ID	ATR1	ATR2	ATR3	VARIABLE OBJETIVO
1				0.2
2				0.9
3				0.5
4				0.1
n				0.4

La variable que queremos predecir es una variable continua, como una probabilidad, edad, valor numérico

Regresión

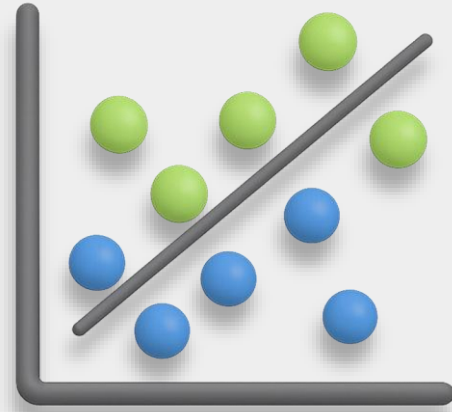
Regresión lineal

Aprendizaje supervisado

Predice **valores numéricos continuos**

- Encuentra relación lineal entre variables
- **Resultado:** línea que mejor ajusta los datos
- **Aplicaciones:** precios, ventas, temperaturas, tiempo

Ejemplo: Predecir precio de casa según metros cuadrados



Clasificación (supervisado)

Las etiquetas son categorías. Ejemplo: enfermo/sano, gato/perro/pájaro, spam/no spam.

ID	ATR1	ATR2	ATR3	VARIABLE OBJETIVO
1				ALTO
2				BAJO
3				MEDIO
4				ALTO
n				MEDIO

La variable que queremos predecir es una clase (categoría)

Clasificación

Aprendizaje no supervisado

Este modelo tiene datos de entrada, pero no se busca una salida en particular. Implica modelar las características de un conjunto de datos sin referencia a ninguna etiqueta.

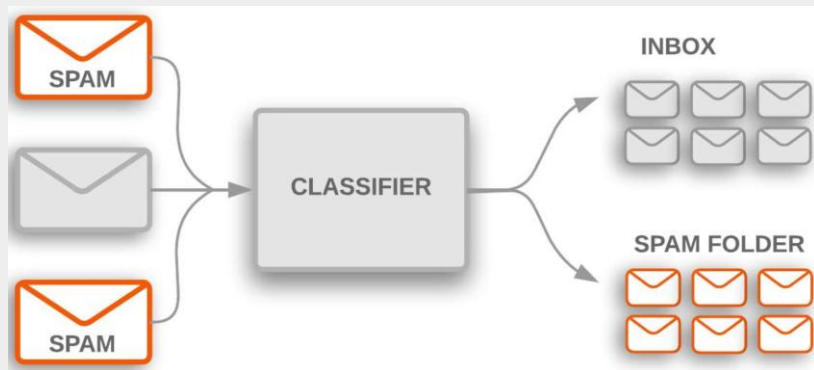
La función de este tipo de algoritmos es encontrar patrones de similitud.

Se utiliza para tareas como **clustering** (agrupación) y **reducción de la dimensionalidad**.



Clasificación binaria

Aprendizaje supervisado



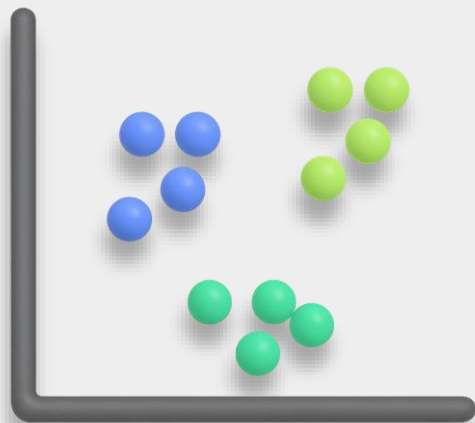
Clasifica **datos en dos categorías**.

- Decisión sí/no, verdadero/falso
- Asigna probabilidad a cada clase
- **Aplicaciones:** diagnósticos, aprobación préstamos

Ejemplo: Email es spam (1) o no spam (0)

Clustering

Aprendizaje no supervisado



Agrupar datos similares sin etiquetas previas.

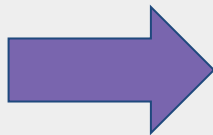
- Encuentra grupos naturales en datos
- No conoce respuesta correcta de antemano
- **Aplicaciones:** segmentación clientes, organización de documentos

Ejemplo: Descubrir tipos de compradores en una tienda

Identifica el algoritmo



- Banco detecta transacciones fraudulentas
- Empresa agrupa productos por características
- App predice tiempo de entrega de pedidos
- Sistema recomienda contenido a usuarios
- Hospital predice días de hospitalización
- Tienda online descubre tipos de compradores



- Clasificación binaria
- Cl
- Regresión
- Clustering
- Regresión
- Clustering



Proceso

1. **Definir problema:** qué queremos predecir
2. **Recolección de datos:** reunir información relevante
3. **Preparación de datos:** limpiar y organizar
4. **Entrenamiento:** el algoritmo aprende patrones
5. **Evaluación:** medir rendimiento con métricas

Modelos y algoritmos

Algoritmo es el conjunto de instrucciones o reglas

VS

Modelo es el resultado del algoritmo entrenado

- Basado en un determinado número de ejemplos en caso de supervisión, o agrupando por similitud
- Se busca generalizar, aprender conceptos a partir de un conjunto de ejemplos y sus características. Cuantos más ejemplos, probablemente sea más fácil la tarea
- Son robustos sistemas de regresión, capaces de ajustarse a una altísima dimensionalidad y una enorme complejidad, difícil de entender.
- El aprendizaje inductivo consiste en construir un modelogeneral a partir de información específica (instancias).

Forma de trabajo

1. **Selección del algoritmo:** se elige por diferentes criterios
2. **Entrenamiento:** usado para que el modelo aprenda patrones. Conforme al algoritmo escogido y los datos que se tienen hay que ver si el entrenamiento da resultados.
3. **Prueba:** Evaluación de la calidad, para determinar el desempeño del algoritmo con datos nuevos (desconocidos). Se utilizan métricas y métodos para decidir si el algoritmo es adecuado o se debe modificar
4. **Validación (opcional):** ajusta hiper parámetros sin tocar el conjunto de prueba. Se los modifica según el tipo de situación, los datos y las métricas arrojadas durante y tras el entrenamiento realizado. Volver al paso 2
5. **Objetivos y Métricas:** Si se está satisfecho, fin de la tarea y modelo entrenado, si no es así, debemos volver al paso 1.

Separar los conjuntos asegura que el modelo generalice bien

Métricas



Son criterios de evaluación que **permiten medir el rendimiento de un modelo** y comprobar qué tan bien realiza sus predicciones.

Métricas de evaluación

- **Accuracy:** Porcentaje de predicciones correctas
- **Matriz de confusión:** tabla que muestra aciertos y errores en predicciones
- **Error promedio:** distancia entre predicciones y valores reales

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision Value $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Se usan después de entrenar el modelo para validar resultados

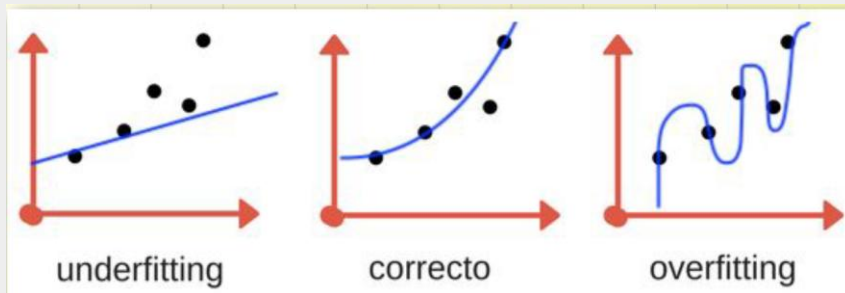
Problemas comunes

Underfitting:

- Modelo demasiado simple, no captura patrones importantes
- Establece una línea divisoria demasiado generalista
- Bajo desempeño para hacer una predicción

Overfitting:

- El modelo está demasiado ajustado a los datos
- Memoriza datos de entrenamiento pero falla con datos nuevos
- Se aprendió los datos de memoria, pero no aprendió a generalizar
- No va a funcionar tan bien para los datos nuevos que no haya visto



Ocurren durante el entrenamiento y se detectan en la evaluación

Venta de ropa online



Tu empresa quiere usar Machine Learning para aumentar ventas.

- ¿Qué problemas de negocio podrías resolver con ML?
- ¿Qué tipo de algoritmo usarías para cada problema?
- ¿Qué limitaciones o riesgos consideras importantes?
- ¿Cómo sabrías si la solución es exitosa?



Proyecto

Tienda Aurelion

- **Documentación:** notebook Markdown
- **Desarrollo técnico:** programa Python
- **Visualización de datos:** dashboard en Power BI
- **Presentación oral:** problema, solución y hallazgos



Diseño conceptual ML

Trabajo en equipo



En relación a la base de datos.

1. Define el **objetivo** (predecir o clasificar)
2. Elige y justifica el **algoritmo**
3. Indica **entradas** (X) y **salida** (y)
4. Especifica las **métricas** de evaluación



Retro

¿Cómo nos vamos?

- ¿Qué fue lo más útil de la clase?
- ¿Qué parte te costó más?
- ¿Qué te gustaría repasar o reforzar?