

Clase 9: Tipos de modelos y ajuste en Machine Learning

FM849 - Programación Científica para Proyectos de Inteligencia Artificial (IA)

9 de enero de 2026

Motivación

- ▶ Hasta ahora, hemos visto lo básico para introducir conceptos relacionados con aplicaciones de inteligencia artificial.
- ▶ Además, hemos visto técnicas que nos permiten entender los datos y prepararlos para las tareas que podamos resolver.

Antes de ir de lleno con la resolución de tareas usando inteligencia artificial, debemos entender **qué es un modelo**, **qué tipos existen**, y **cuál es el concepto de “ajuste”** en este contexto.

¿Qué es un modelo?

Bajo el contexto de inteligencia artificial, un modelo lo podemos pensar, de forma muy simplificada, como un mapa que nos permite movernos en una ciudad. ¿Por qué?

- ▶ Sabemos que la ciudad (*realidad*) es compleja, y es casi imposible conocerla en su totalidad (*capturar todas sus variables*).
- ▶ Para trasladarnos de un punto a otro, no necesitamos una representación perfecta (*conocer todas las variables*), sino una aproximación que nos permita entender cómo llegar a un destino. En este caso, dicha aproximación es el mapa (*modelo*), donde se imprime lo justo y necesario (*variables que se poseen*).

Es importante hacer énfasis en que el mapa (*modelo*) se construye a partir de información previa (*datos aprendidos*). Esa información, si es representativa de la ciudad (*realidad*), deberían permitir a cualquier persona que nunca la haya visitado (*datos nuevos*) trasladarse sin problemas.

Datos de entrenamiento y datos nuevos

Antes de pasar a los tipos de modelos en inteligencia artificial, debemos entender que los datos se particionan en dos (y a veces tres) conjuntos. Por ahora, veremos los dos conjuntos que siempre están presentes:

- ▶ **Conjunto de entrenamiento:** permite alimentar al modelo de conocimiento. Aquí es donde el modelo aprende patrones y relaciones entre las variables.
- ▶ **Conjunto de prueba** (datos nuevos): permite evaluar el desempeño del modelo con datos no vistos durante el entrenamiento, para ver si existe un buen poder de generalización.

Esto permite monitorear que el modelo no memorice patrones triviales, sino que capture comportamientos complejos que se aproximen mejor a la realidad.

Modelo de aprendizaje supervisado

Es el tipo de aprendizaje más común en Machine Learning. Se llama “supervisado” porque los datos de entrenamiento incluyen las respuestas correctas (*etiquetas*, denotadas por y) para las variables que **no** sean el objetivo de predicción (denotadas por X).

Ejemplo

Si queremos predecir el precio de una casa, el conjunto de entrenamiento incluirá variables de las casas (tamaño, ubicación, número de habitaciones, etc.) que no son el objetivo de predicción (X) junto con sus precios reales (y).

El modelo aprenderá a asociar estas características con los precios para hacer predicciones sobre los precios de nuevas casas.

Aplicaciones de aprendizaje supervisado

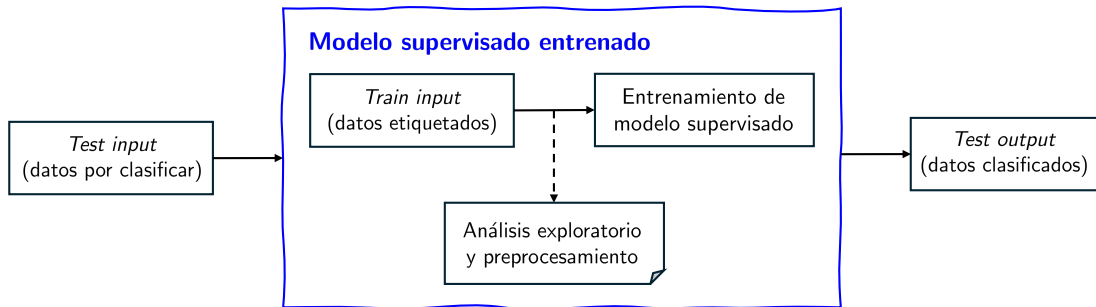
El aprendizaje supervisado se utiliza en una amplia variedad de aplicaciones, tales como:

- ▶ Clasificación de correos electrónicos como spam o no spam.
- ▶ Diagnóstico médico basado en síntomas y resultados de pruebas.
- ▶ Reconocimiento de dígitos manuscritos (ver la demostración de Yann LeCun en 1989 en [YouTube](#)).
- ▶ Pronósticos de temperaturas y condiciones climáticas.

Y estos son sólo algunos de los ejemplos. Obviamente hay muchos más que pueden ser interesantes.

Diagrama de aprendizaje supervisado

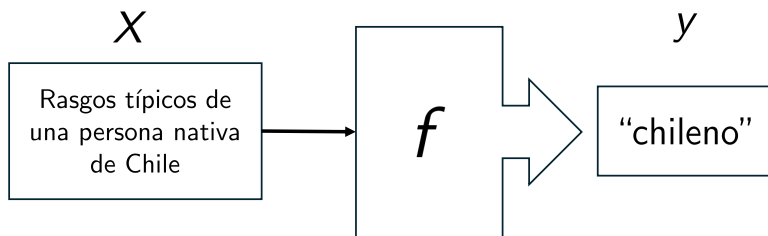
De forma simplificada, podemos pensar el funcionamiento del aprendizaje supervisado como se muestra en el siguiente diagrama:



Formalizando el aprendizaje supervisado

Formalmente, lo que se busca es aproximar una función f que modela la relación entre las variables predictoras X y las etiquetas y .

Piensen que la función f es la generadora de las etiquetas reales, es decir, $y = f(X)$.



Formalizando el aprendizaje supervisado

Como mencionamos al inicio, en la mayoría de los casos, es imposible aprender la forma exacta de la función f que genera la relación $y = f(X)$, porque nunca tendremos acceso a todos los datos. Buscaremos una aproximación $\hat{y} = \hat{f}(X)$ que funcione “suficientemente bien”.

“Suficientemente bien” significa minimizar un criterio de error o discrepancia entre las etiquetas reales y , y las predicciones $\hat{y} = \hat{f}(X)$.

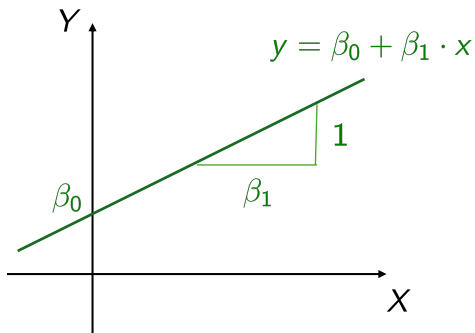
Ejemplo

Si consideramos que la discrepancia se mide como el número de errores de clasificación, entonces un buen modelo es aquel que se equivoca lo menos posible. Esta métrica se conoce como *zero-one loss* (E):

$$E = \sum_{i=1}^n [y_i \neq \hat{y}_i]$$

Formalizando el aprendizaje supervisado

El aprendizaje de la función \hat{f} se realiza con los datos de entrenamiento. Piensen, por ejemplo, que hay un fenómeno que se puede modelar con una recta $y = \beta_0 + \beta_1 \cdot x = f(x)$. Esta recta se muestra en la siguiente imagen:



Nuestro problema es justamente el desconocimiento sobre los coeficientes β_0 y β_1 .

Si probamos entrenando un modelo lineal, el modelo nos entregará una aproximación $\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$ que debiese parecerse a la recta real. Esto **no necesariamente es así**, dado que el modelo sólo aprendió de los datos de entrenamiento, y puede tener poca generalización.

Tipos de ajuste: *underfit* y *overfit*

Cuando hablamos de poder de generalización, nos referimos a qué tan bien el modelo se adapta a nuevos datos que nunca vio en su entrenamiento.

Una buena analogía para un modelo consiste en un estudiante que debe estudiar para una prueba.

- ▶ Si el estudiante no estudia, o estudia muy poco, es probable que ni siquiera haya aprendido lo suficiente para resolver preguntas básicas. Esto se conoce como *underfitting* (subajuste).
- ▶ Si el estudiante estudia sólo de las pruebas pasadas, es probable que memorice las respuestas de esas pruebas sin entender los conceptos de manera general. Esto se conoce como *overfitting* (sobreajuste).

El objetivo es encontrar un equilibrio, donde el estudiante entienda los conceptos y pueda aplicarlos a nuevas preguntas que no ha visto antes.

Modelo de aprendizaje no supervisado

Ya que conocemos el aprendizaje supervisado, podemos definir el aprendizaje no supervisado como aquel donde los datos de entrenamiento no incluyen etiquetas o respuestas correctas.

En este caso, el modelo debe encontrar patrones, estructuras o relaciones en los datos por sí mismo, sin una guía explícita.

Ejemplo

Si tenemos un conjunto de datos con características de clientes (edad, ingresos, hábitos de compra, etc.) pero sin etiquetas que indiquen segmentos específicos, un modelo de aprendizaje no supervisado podría identificar grupos de clientes con comportamientos similares. Esto puede permitir generar recomendaciones dirigidas.

Aplicaciones de aprendizaje no supervisado

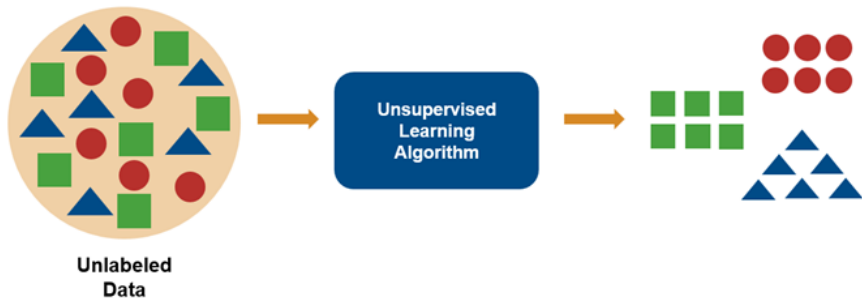
El aprendizaje no supervisado se utiliza en diversas aplicaciones, tales como:

- ▶ Segmentación de clientes en marketing (p. ej., Amazon, Mercado Libre, etc.).
- ▶ Agrupamiento de documentos similares en sistemas de recomendación (p. ej., categorización en diarios según el texto de las noticias).
- ▶ Detección de fraude bancario mediante el reconocimiento de anomalías.

Al igual que con el aprendizaje supervisado, estos son sólo algunos ejemplos representativos.


¿Cómo se ve el aprendizaje no supervisado?

La siguiente imagen ilustra una de las tareas que se pueden realizar con aprendizaje no supervisado: el agrupamiento (*clustering*) de datos en diferentes grupos basados en sus características.



Otros tipos de aprendizaje

Existen también otros tipos de aprendizaje que no veremos en el curso, pero los mencionamos para quienes estén interesados en seguir aprendiendo:

- ▶ Aprendizaje por refuerzo: el modelo aprende a tomar decisiones mediante la interacción con un entorno, recibiendo recompensas o castigos según sus acciones. Ejemplo:  *speedruns* en el Super Mario Bros. original.
- ▶ Aprendizaje semisupervisado: combina datos etiquetados y no etiquetados para mejorar el aprendizaje del modelo. Ejemplo: Google Fotos.
- ▶ Aprendizaje autosupervisado: el modelo genera sus propias etiquetas a partir de los datos no etiquetados, permitiendo aprender representaciones útiles. Ejemplo: procesamiento de lenguaje natural (p. ej., GPT).
- ▶ Aprendizaje generativo: el modelo aprende a generar nuevos datos similares a los datos de entrenamiento. Ejemplo: generación de imágenes realistas (p. ej., Dall-E).