

Extracción del Conocimiento en Bases de Datos

Unidad III. Análisis supervisado

Objetivo de la unidad

El alumno implementará algoritmos de análisis supervisado para aplicarlos en la predicción y clasificación de nuevas entradas de datos.

Temas

- Algoritmos de aprendizaje supervisado.
- Evaluación de modelos de aprendizaje supervisado.

ANÁLISIS SUPERVISADO, APLICACIONES Y CLASIFICACIÓN



APRENDIZAJE SUPERVISADO

Es una rama de Machine Learning , un método de análisis de datos que utiliza algoritmos que aprenden iterativamente de los datos para permitir que los ordenadores encuentren información escondida sin tener que programar de manera explícita dónde buscar.

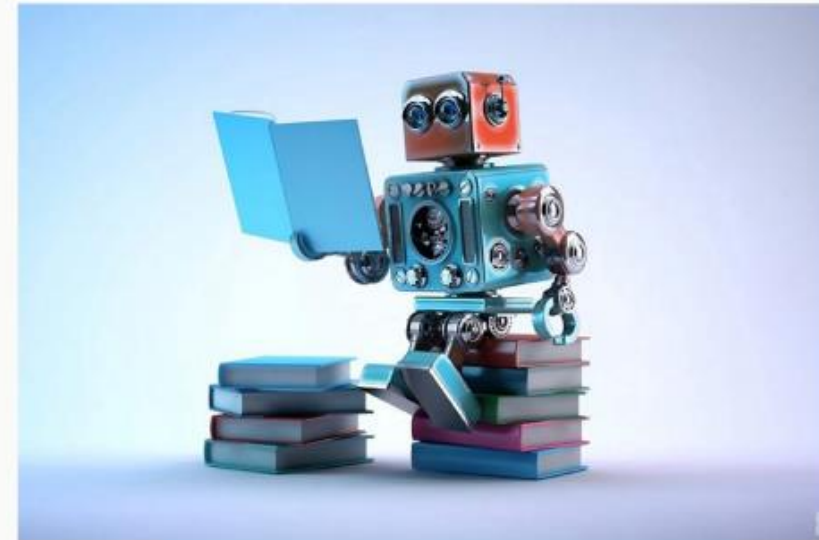


Trabaja con datos etiquetados, para los que ya conoce la respuesta de destino.

En base a un histórico, trata de buscar patrones.

Aprende a asignar una etiqueta de salida o función.

Se utiliza en dos tipos de problemas: clasificación y regresión.



PASOS BÁSICOS DEL APRENDIZAJE SUPERVISADO

1

SELECCIÓN

Seleccionar el tipo de datos de entrenamiento

2

FILTRAR DATOS

Recopilar y limpiar los datos de entrenamiento

3

MODELO

Elegir un modelo utilizando un algoritmo de aprendizaje supervisado

4

ENTRENAMIENTO

Entrenar el modelo.

5

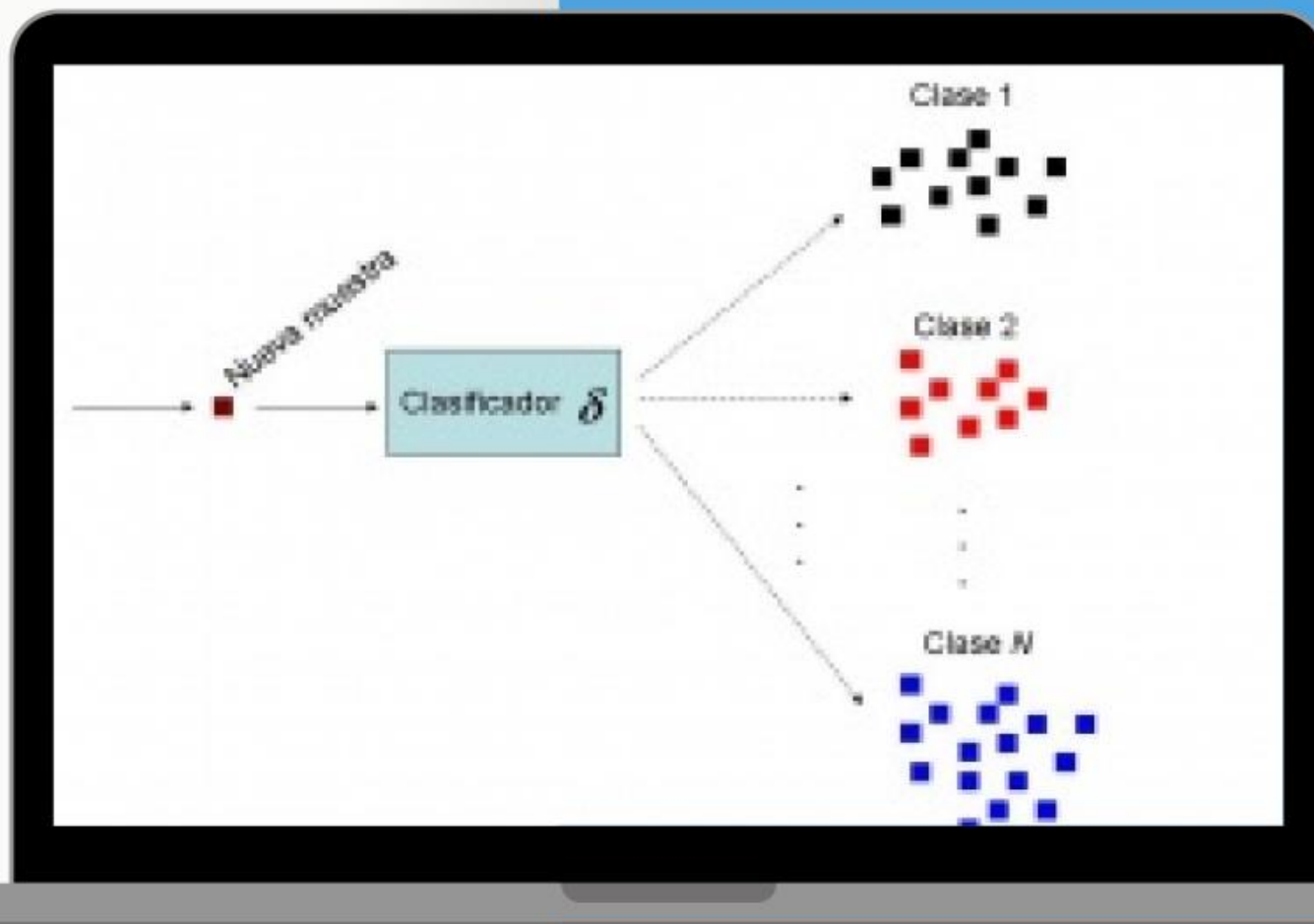
EVALUACIÓN

Realizar predicciones y evaluar el modelo.

CLASIFICACIÓN

Es el lugar donde se entrena a un algoritmo para clasificar los datos de entrada en variables discretas.

Durante el entrenamiento, los algoritmos reciben datos de entrada de entrenamiento con una etiqueta de "clasificación".



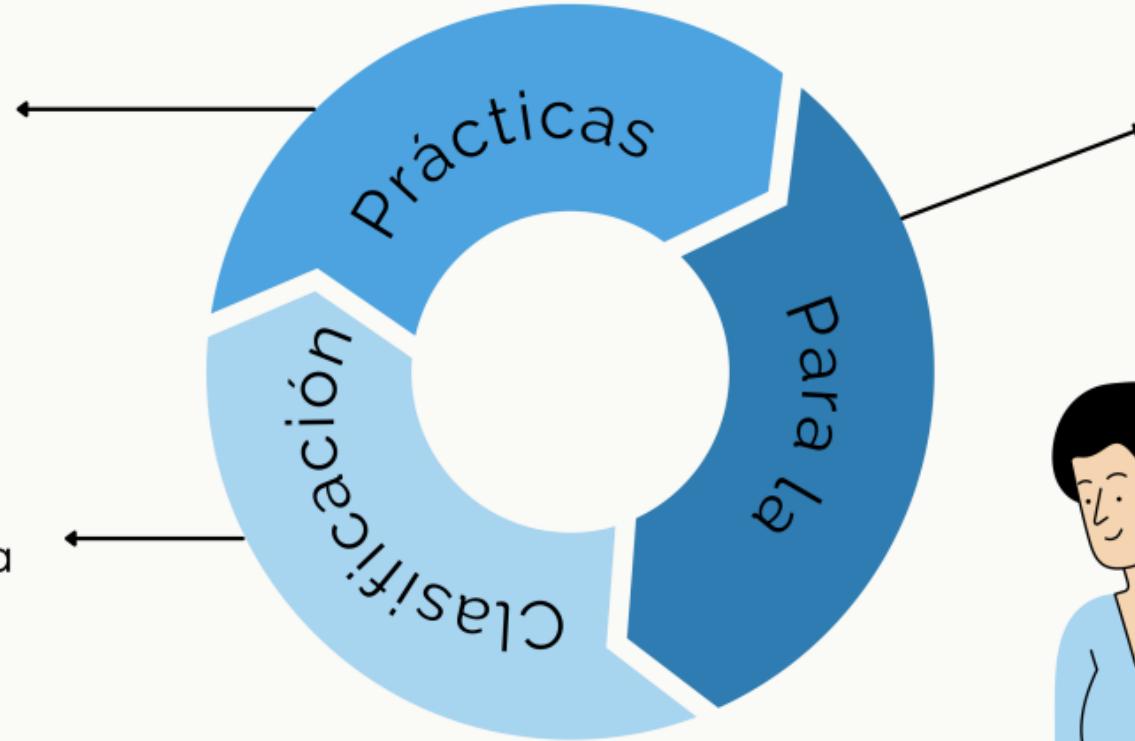
APLICACIONES

Clasificación binaria:
Clasifica los datos de
entrada en uno de dos
grupos posibles.

- Detección de Spam
- Predicción de migración de clientes
- Predicción de conversión

Clasificación multiclase:
El conjunto de datos de
entrenamiento se etiqueta
con una de las múltiples
clases posibles.

- Clasificación de rostros
- Clasificación de correo electrónico
- Bosques aleatorios



Clasificación de etiquetas
múltiples:

Significa que los mismos
datos de entrada podrían
clasificarse en diferentes
compartimientos

- Detección de fotos
- Clasificación de audio/video
- Clasificación de documentos



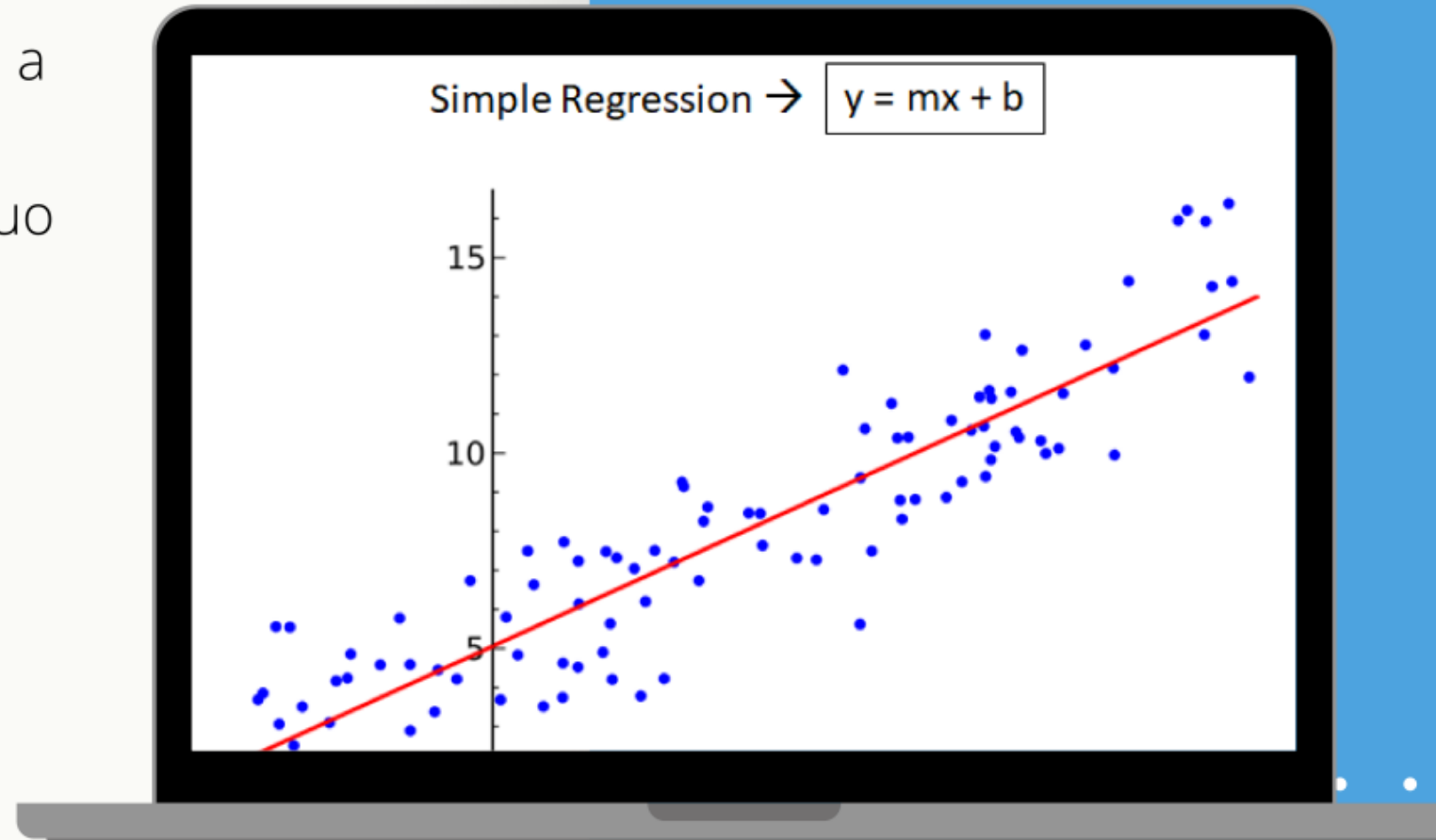
DEFINICION

El análisis de regresión es un subcampo del aprendizaje automático supervisado cuyo objetivo es establecer un método para la relación entre un cierto número de características y una variable objetivo continua.



REGRESIÓN

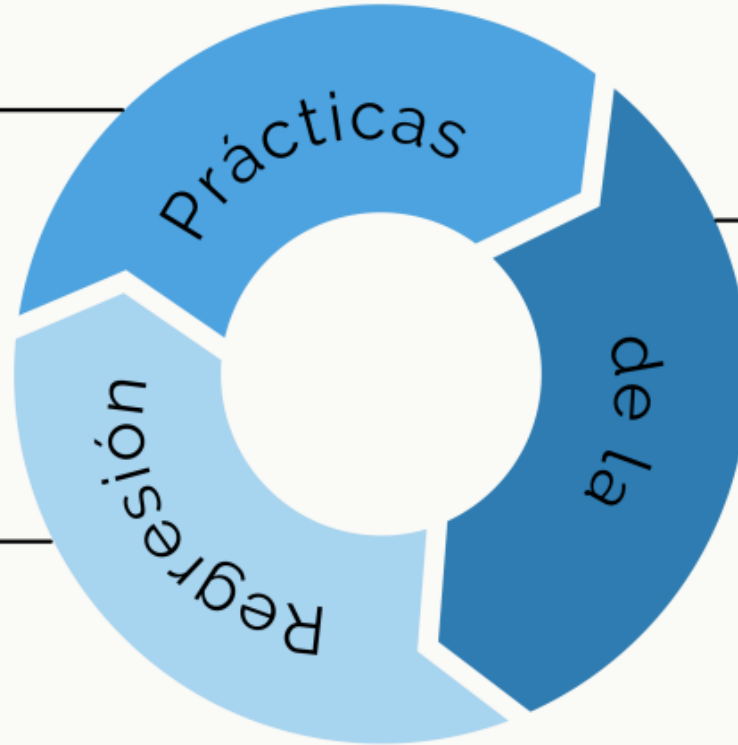
Es un método de aprendizaje supervisado en el que se entrena a un algoritmo para predecir una salida a partir de un rango continuo de valores posibles.



APLICACIONES

Regresión lineal:
Entrena a un algoritmo
para encontrar una
relación lineal entre los
datos de entrada y salida.

Regresión logística:
Se utiliza para
determinar la
probabilidad de que
ocurra un evento.

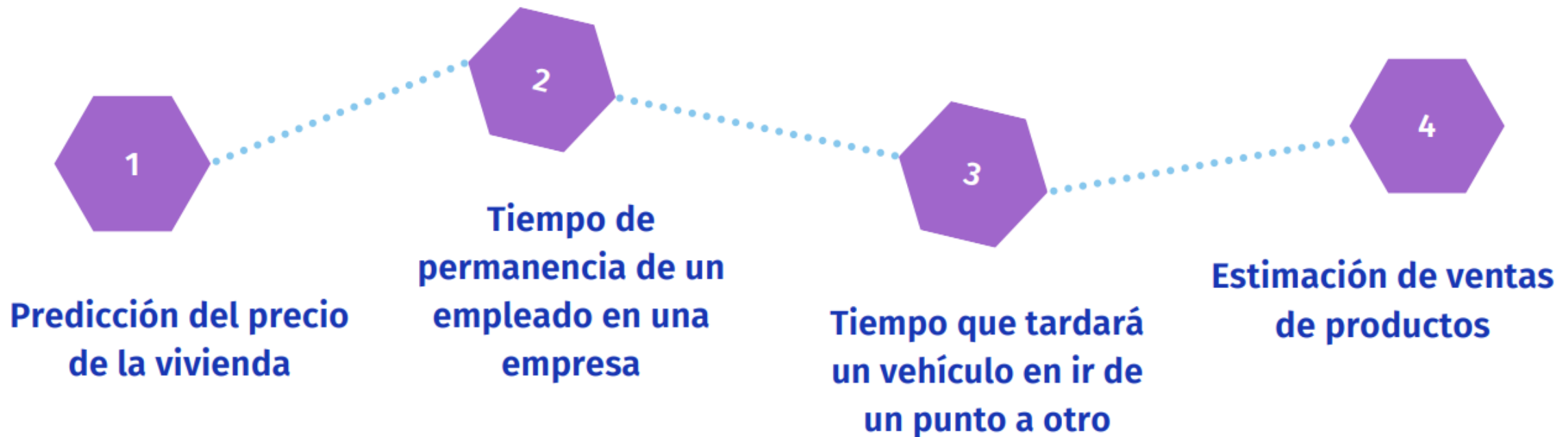


Regresión polinomial:
Se utiliza para un
conjunto de datos
más complejo que no
encajaría en una
regresión lineal.





Algunos de los usos son los siguientes





UT DE LA COSTA

Técnicas de Machine Learning para regresión

Regresión lineal

Regresión logística

Árboles de decisión

Deep Learning.

Regresion Lineal

Es un algoritmo de aprendizaje supervisado que es utilizado en machine learning y en estadística, en su versión más sencilla lo que se hace es solo dibujar una línea recta. y en estadística sería lo mismo solo sería hacer una línea de el punto Y al punto X



Regresion Logística

Es un método estadístico que se usa para resolver problemas de clasificación binario, donde el resultado solo puede ser naturaleza dicotómica, o sea que solo puede tomar dos valores posibles



Arboles de decisión

Esto nos permite la construcción de modelos predictivos de analítica de datos para el big data basado en su clasificación según ciertas características o propiedades, o en la regresión mediante la relación entre distintas variables para predecir el valor de otra



Deep Learning

En lugar de enseñarle a ordenador una lista enorme de reglas para solventar un problema, le damos un modelo que pueda evaluar ejemplos y una pequeña colección de instrucciones para modificar el modelo cuando se produzcan errores. Con el tiempo esperamos que esos modelos sean capaces de solucionar el problema de forma extremadamente precisa, gracias a que el sistema es capaz de extraer patrones.

