EDEM
Escuela de Empresarios

EDEM Escuela de Empresarios

Machine Learning 0 - Intro

horus·ml

Jesús Prada Alonso - HORUS ML

Curso Máster en Data Analytics - Edición 4

Fecha 30/03/2023

RESUMEN FINAL

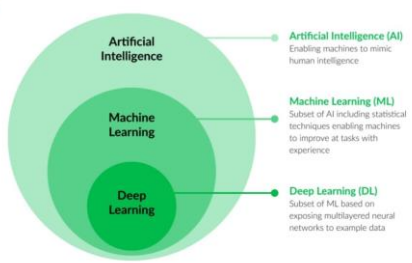
EDEM
Escuela de Empresarios

RESUMEN

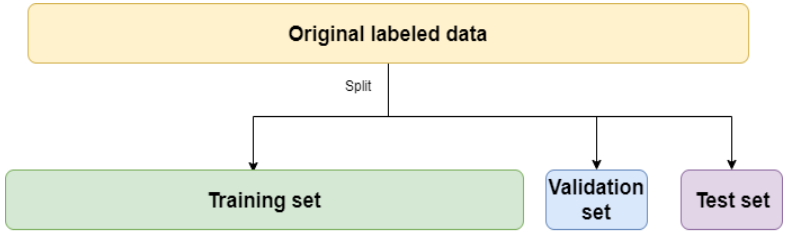
EDEM
Escuela de Empresarios

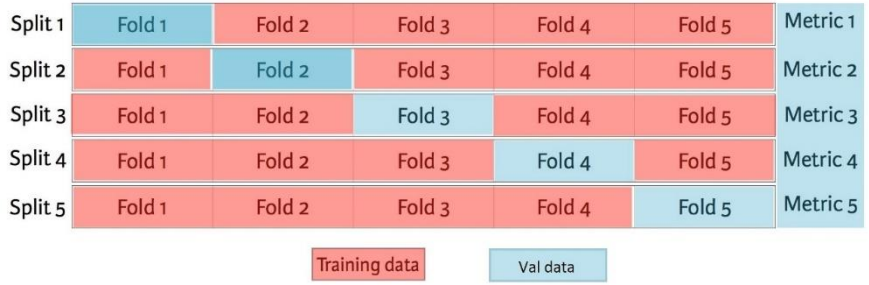
Definición Machine Learning

***Machine Learning**, ML, o **Aprendizaje Automático** es una rama de la Inteligencia Artificial cuyo objetivo es construir sistemas que aprendan automáticamente de los datos.*



Cross Validation y validación fija





- TRAINING : Datos de los que los modelos extraen patrones.
- VALIDATION : Se emplea para seleccionar el mejor de los modelos entrenados en metamodelización.
- TEST : Proporciona el error real esperado con el modelo seleccionado.

Grid Search

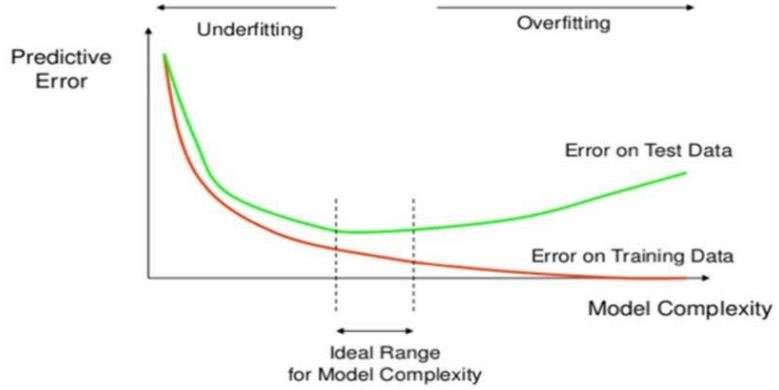
- Los modelos ML suelen incluir un conjunto de **hiperparámetros** que nos permiten controlar su comportamiento.
- De su correcta elección dependerá la bondad del modelo entrenado.

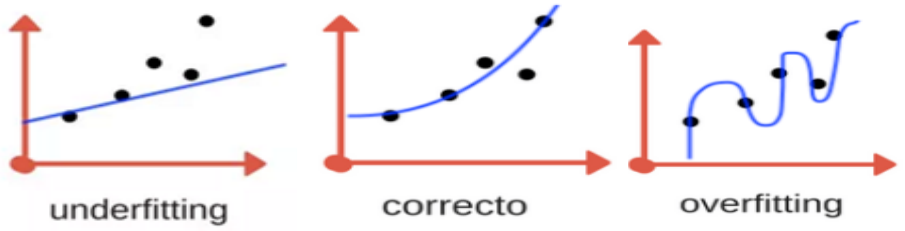
par1/par2	10	100	1000
0.1	0.3	0.22	0.25
0.01	0.15	0.14	0.14
0.001	0.35	0.05	0.11

RESUMEN

EDEM
Escuela de Empresarios

Overfitting y Underfitting





¿Cómo detectar el overfitting?

Validación tiene un error mucho mayor que en train

¿Cómo detectar el underfitting?

El error de train parece demasiado elevado o da la misma respuesta siempre.

Supervisado VS no Supervisado

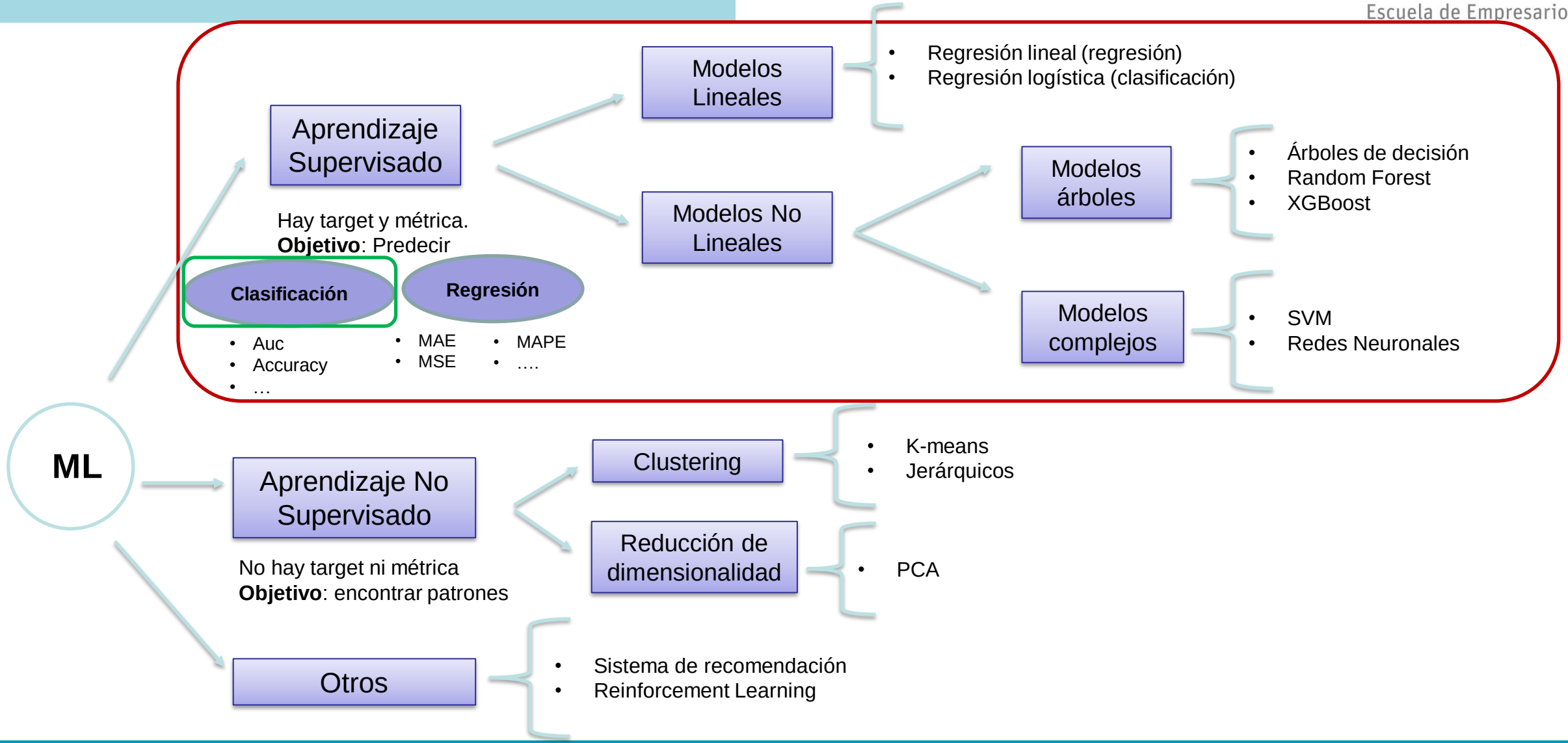
	Supervisado	No supervisado
Etiquetas	SI	NO
Objetivo	Dar predicciones a futuro sobre el conjunto de test	Encontrar patrones en los datos o reducir dimensiones
Modelos	Regresión lineal, árboles, SVM, Redes Neuronales	Clustering, PCA
Ejemplo	Predecir si una transacción es fraudulenta	Encontrar clientes con perfiles similares

Clasificación VS Regresión

	Clasificación	Regresión
Etiquetas	Categóricas.	Numéricas.
Ejemplo	Una imagen es un gato (1) o no (0). 	Precio de alquiler de una casa 
Métrica	AUC	MSE

RESUMEN MODELOS

EDEM
Escuela de Empresarios



RESUMEN MODELOS

EDEM
Escuela de Empresarios

Modelo	Alcance	Definición	Ventaja	Desventaja	Hiperparámetros
Regresión logística	Clasificación	Estima probabilidades utilizando una función logística al resultado de una regresión lineal.	Opción más sencilla para clasificación.	Demasiado sencillo para la mayoría de problemas.	• Lasso o Ridge. • alpha (regularización).
SVM	Ambos	Encontrar el hiperplano de máximo margen o ajuste.	Uno de los modelos más potentes. Solución asegurada y única.	Entrenamiento costoso. Sensible a hiperparámetros.	• C • γ • ε
Árbol de decisión	Ambos	Cada nodo es una variable, cada rama una decisión y cada nodo hoja un output.	Interpretable. Admite variables categóricas.	Demasiado sencillo para la mayoría de problemas.	max depth, min samples split, min samples leaf, max features.
Random Forest	Ambos	Combinación de árboles de decisión con voto.	Proporciona variable importance. Robusto a hiperparámetros.	Menor potencial predictivo que SVM, NN o XGB.	• ntree • mtry • nodesize
XGBoost	Ambos	Combinación iterativa de árboles o RF que corrigen errores de los anteriores.	Gran potencial predictivo. Muy rápido.	Muchos hiperparámetros que optimizar.	nrounds, eta, gamma, max depth, min child weight, subsample, colsample bytree, num parallel tree, lambda, alpha, early_stopping_rounds.

EDEM
Escuela de Empresarios

EDEM Escuela de Empresarios

Jesús Prada Alonso
jesus.prada@horusml.com