Tipos de algoritmos:

* Aprendizaje supervisado: algoritmos de predicción y de clasificación
* Aprendizaje no supervisado: clusterización y reducción de dimensionalidad
* Aprendizaje por refuerzo

A partir de los datos de base se divide la base en un conjunto de datos de entrenamiento y un conjunto de datos de test.

Conjunto de entrenamiento un 80% de los datos y un 20% para el conjunto de test

test\_size =0.2 🡪 equivale a que me recoja del conjunto el 20%

model.fit(x,y) 🡪 va entrenando el modelo

model.predict (x) -🡪 predecir el valor de y a partir de la x

lin.reg.score (x,y) \*100 🡪 nos da el porcentaje de aciertos en la predicción . Si este porcentaje es del 100% el MAE es 0. No es una buena métrica de los errores para una regresión

Nuestro modelo partirá de pruebas con el conjunto de datos de entrenamiento

El conjunto de test nos permitirá ver cómo ha aprendido el modelo. El criterio de fiabilidad y validez del modelo lo establece el científico de datos en función del modelo

Conjuntos de validación son subconjuntos del conjunto de entrenamiento para ir validando el modelo antes de aplicarlo al conjunto de tests

Machine learning.- detectar patrones y hacer predicciones a partir de estos patrones

No supervisado.- No hay una variable a predecir ni tengo los valores etiquetados. Para hacer clúster y detectar anomalías

Machine learning (aprendizaje automático)

Deep learnint es un algoritmo de ML pero que utiliza muchas capas y cada una de ellas hace una acción determinada

Inteligencia artificial. Parte del ML con el objetivo de imitar comportamientos cognitivos del ser humano. Chat box por ej.

Regresión lineal

Variable dependiente la llaman variable regresando

Variable independiente la llaman variable regresor

Y = a + bx

“a” es la secante que es el punto donde la recta corta con el eje y. Este valor en la recta es el valor de y cuando x vale 0

“b”: cuando se incrementa el eje x cuanto se incrementa un punto el eje y. Si el eje x aumenta un punto y el eje y dos la pendiente es 2

<0 las dos crecen

<0 una crece y otra decrece

0: no hay asociación lineal

MAE.- media de errores absolutos. Nos da el error medio con el que nos estamos equivocando al hacer una predicción

MSE.-Error cuadrado medio

RMSE.- Cuanto más mayor sea que el MAE, más outliers hay

Lo que se busca siempre es calcular la recta con el menor error

Todos los algoritmos tienden a encontrar el mínimo error de una función

Np.random.seed(1111).- la semilla permite que si otra persona genera otros números aleatorios los que se generan son los mismos números aleatorios

Para generar una muestra aleatoria: np.random.sample(100)

Estos números aleatorios de muestra generados, como partimos de la misma semilla serán los mismos que han salido en clase

Con random sample no se repiten números en la muestra generada

X= np.arange(100) genera una lista de 100 números

x = x.reshape(-1,1) 🡪 pasa un vector a matriz colocando en una columna los datos del vector. Transpone la fila a columna

x.shape

(100,1)

Cuando hay gran cantidad de datos mejor utilizar el modelo de ir por partes para entrenar el modelo. Ir iterando paso a paso más que darle directamente:

new\_model2 = LinearRegression()

new\_model2.fit(x,y)