

MACHINE LEARNING CON SCIKIT-LEARN

Abraham Jain Jiménez

Learning Machines

Una *learning machine* es un objeto Θ que puede ser conceptualizado como una función que mapea una base de datos D_f a un estimador \hat{f} de la función original f :

$$\Theta : D_f \rightarrow \hat{f}$$

El estimador \hat{f} es producido por el modelo de Machine Learning implementado para el problema.

scikit-learn

Scikit-learn es una biblioteca de código abierto para aprendizaje automático que admite aprendizaje supervisado y no supervisado. También proporciona diversas herramientas para el ajuste de modelos, el preprocesamiento de datos, la selección de modelos, la evaluación de modelos y muchas otras utilidades.

Está construida sobre las bibliotecas NumPy, SciPy y matplotlib.

Machine Learning (ML) con scikit-learn

Requisitos que deben cumplir las bases de datos para implementar modelos de ML sobre ellas con **scikit-learn**:

- No tener *missing values*.
- Formato numérico.
- Almacenadas en un **DataFrame** de **Pandas** o un arreglo **NumPy**.

Es por ello la importancia de efectuar un EDA (*Exploratory Data Analysis*) sobre las bases de datos.