

*Práctica 2*

Aplicación de métodos de  
*Machine Learning* a datos  
biomédicos

**Introducción a la Bioingeniería**

Grado en Ingeniería en Sistemas de Telecomunicación

Luis Bote Curiel

# Contenido

1. Introducción
2. Métodos de regresión
3. Métodos de clasificación
4. Métodos de clustering
5. Librerías de ML en Python
6. Práctica

# Introducción

- Tipos de métodos de *Machine Learning* (ML):
  - **Supervisados**: son aquellos en los que se tienen variables de entrada (X) y una variable de salida (Y) y se utiliza un algoritmo para aprender la función de asignación de la entrada a la salida.

$$Y = f(X)$$

- Clasificación: La variable de salida es una categoría.
- Regresión: La variable de salida es un número real.

# Introducción

- Tipos de métodos de *Machine Learning* (ML):
  - **No supervisados**: son aquellos en los que sólo se dispone de datos de entrada (X) y ninguna variable de salida, y cuyo objetivo es modelar la estructura o distribución subyacente de los datos para aprender más sobre ellos.
    - Se denominan también algoritmos de **clustering**.

# Métodos de regresión

- Algunos métodos para **regresión** son:
  - Linear Regression
    - LASSO Regression
    - Ridge Regression
  - Decision Tree Regression
  - Random Forest Regression
  - Support Vector Regression
  - K-nearest neighbors (KNN) for regression
  - Neural Networks for regression
  - Extreme Gradient Boosting (XGBoost) [Gradient boosting algorithm] for regression

# Métodos de regresión

- Algunas métricas para métodos de **regresión** son:
  - Mean Squared Error
  - Root Mean Squared Error
  - Mean Absolute Error

# Métodos de clasificación

- Algunos métodos para **clasificación** son:
  - Logistic Regression
  - Decision Tree
  - Random Forest
  - Support Vector Machine
  - K-nearest neighbors (KNN)
  - Naïve Bayes
  - Neural Networks
  - Extreme Gradient Boosting (XGBoost) [Gradient boosting algorithm]

# Métodos de clasificación

- Algunas métricas para métodos de **clasificación** son:
  - Accuracy
  - Confusion matrix
  - F1 score
  - ROC curve
  - AUC



# Métodos de clasificación

- Algunos métodos para **clustering** son:
  - K-Means
  - Mean-Shift Clustering
  - Mixture of Gaussians

# Librerías de ML en Python

- Scikit-Learn: <https://scikit-learn.org>
- XGBoost: <https://xgboost.readthedocs.io/>
- PyTorch: <https://pytorch.org>
- TensorFlow: <https://www.tensorflow.org>

# Práctica

## **Aplicación de métodos de *Machine Learning* a datos biomédicos**

- La práctica consistirá en aplicar métodos de ML a datos biomédicos usando librerías de Python. En concreto, el alumno tendrá que seleccionar un mínimo de 2 métodos de regresión y 2 métodos de clasificación. Los métodos de regresión se utilizarán en el dataset *peruvian\_blood\_pressures.csv* para predecir la variable *systol* y la variable *diastol*. Los métodos de clasificación se utilizarán en el dataset *south\_africa\_chd.csv* para predecir la variable *chd*. Tanto en los métodos de regresión como en los métodos de clasificación se utilizarán, como mínimo, 2 métricas.

# Bibliografía

- Geron, Aurelien (3 Edition). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow
- Raschka, Sebastian (2 Edition). Python Machine Learning
- An introduction to machine learning with scikit-learn (<https://scikit-learn.org/stable/tutorial/basic/tutorial.html>)
- Scikit-learn course (<https://inria.github.io/scikit-learn-mooc>)
- Get Started with XGBoost ([https://xgboost.readthedocs.io/en/stable/get\\_started.html](https://xgboost.readthedocs.io/en/stable/get_started.html))
- Deep Learning with PyTorch: A 60 Minute Blitz ([https://pytorch.org/tutorials/beginner/deep\\_learning\\_60min\\_blitz.html](https://pytorch.org/tutorials/beginner/deep_learning_60min_blitz.html))