



Universidad Francisco Marroquín
Facultad de Ciencias Económicas
Fundamentals of Machine Learning
04 de marzo de 2020

Laboratorio 3

Regresión Logística

Instrucciones: Usted usará Python a través de un Jupyter Notebook para llevar a cabo este laboratorio. Recuerde utilizar comentarios para describir lo que está haciendo en cada parte del proceso y así dejar constancia del entendimiento de su trabajo. Estará usando el dataset proporcionado dentro del portal. Al finalizar recuerde subir al portal un link a su repositorio en el que se pueda correr su notebook.

El dataset proporcionado es parte de la plataforma Kaggle, dentro del cual se muestran diferentes atributos que pueden servir para identificar si un paciente tiene o no una enfermedad del corazón, como la edad, dolor de pecho, presión sanguínea entre otros, además se muestra si cada uno de los pacientes tiene o no dicho padecimiento (columna nombrada como *target*).

Para este ejercicio se le pide que proporcione un modelo de **regresión logística** que prediga si los pacientes tienen o no un padecimiento cardíaco usando todas las features que considere necesarias. A continuación se mencionan las generalidades de los pasos sugeridos a realizar.

1. Leer el archivo CSV proporcionado (heart.csv, <https://www.kaggle.com/ronitf/heart-disease-uci>) y almacenarlo en un **np.array** para ser trabajado en el notebook.
2. Ajustar un modelo de regresión logística en base al dataset cargado que relacione cualquier subconjunto propio de las variables de los indicadores seleccionados con la probabilidad de tener una enfermedad cardíaca.
3. Asegúrese de hacer los ajustes y trabajos necesarios al modelo que ha implementado para describir y predecir la información sobre la nube de datos.
4. Haga un análisis sobre sus hallazgos, donde mencione claramente las razones por las que considera que su modelo es bueno justificando adecuadamente.