

# Análisis de datos con Python

## Ejercicios Clase 5 - UNSAM 2018

### Regresión Lineal y Logística

1. Modificar el script `LinearRegressionDiabetes.py` (ejemplo del manual) de forma de
  - (a) Ajustar una regresión utilizando todas las (10) variables.
  - (b) Ajustar una regresión utilizando las (8) variables comprendidas entre la segunda y la anteúltima.
  - (c) ¿Cuál de estas dos regresiones ajusta mejor? ¿Qué modelo prefiere entre estos dos?
2. Usando el dataset **Tips** (clase 3)
  - (a) Ajustar una regresión lineal que modele la propina en función del precio de la comida, utilizando todos los datos.
  - (b) Idem, pero utilizando solamente los datos de mujeres no fumadoras. ¿Qué modelo ajusta mejor?
  - (c) Utilizar la regresión anterior para enstimar la propina que deja una mujer no fumadora si la cuenta es de pesos \$20.
3. Usando el Dataset **house\_prices**
  - (a) Realizar un regresion lineal que permita estimar el precio de venta de la casa ('SalePrice') en base a la superficie total ('totSF') de la misma. Realizar un gráfico donde figuren los datos originales y la recta hallada. ¿Ha encontrado una buena aproximacion lineal?
  - (b) Realice una limpieza de los datos y repita el item anterior.

Logística Usar el dataset de **Breast Cancer** (clase 3) para ajustar el diagnóstico con una regresión logística utilizando una, otra y ambas variables `radius_mean` `concavity_mean`. ¿Cómo ajusta cada modelo?