

IMPLEMENTADO MACHINE LEARNING

Práctica 2 – Regresión lineal

Material de lectura:

- Andreas C. Müller, Sarah Guido. Introduction to Machine Learning with Python. Capítulo 2.
- Análisis Descenso de gradiente. <https://towardsdatascience.com/gradient-descent-algorithm-a-deep-dive-cf04e8115f21>
- Linear Regression for Machine Learning. <https://machinelearningmastery.com/linear-regression-for-machine-learning/>
- Scikit Learn: <https://scikit-learn.org/>

- 1- **Introducción al Descenso de gradiente.** El archivo *regresion_lineal_entrenamiento.py* posee un código que ejemplifica el entrenamiento de un modelo lineal con la técnica descenso de gradiente.
 - a. Modificar el parámetro de aprendizaje: Alfa
 - b. Modificar los valores iniciales de m y b.
 - c. Cambie la variable “NORMALIZAR” de 1 a 0 para observar el efecto de entrenar el modelo con las variables con y sin normalización.
- 2- **Boston.** El archivo *regresion_lineal_boston_sklearn.py* presenta un pequeño script donde se entrena un modelo lineal para predecir los precios de las casas según las características del barrio (dataset *Boston.csv*). Analice los coeficientes generados con y sin normalización.
- 3- **Vinos.** Utilice los datasets *winequality-red.csv* y *winequality-white.csv* para predecir la calidad de los vinos (establecida en el atributo “quality”).
 - a. Compare el Error Cuadrático Medio y el Error Absoluto Medio.
 - b. Analice qué ocurre con el error para distintos rangos de calidades de vino. ¿Por qué el error es menor para calidades cercanas al 6?
- 4- **Clima.** El dataset en *wheater.csv* contiene información sobre el clima. Cada registro es la medición de un día distinto.
 - a. Limpie el dataset para tener una vista minable.
 - b. Utilice solo las variables 'Rainfall', 'WindGustSpeed', 'WindSpeed9am', 'WindSpeed3pm', 'Humidity9am', 'Humidity3pm', 'Pressure9am', 'Pressure3pm', 'Temp9am', 'Temp3pm'.
 - c. Intente predecir las diferentes variables de forma individual. Para esto quite la variable a predecir del dataset (x) y utilícela como variable a estimar (y). Por ejemplo:

```
x= data_frame.drop(' Temp3pm', axis=1)
y= data_frame [' Temp3pm']
```
 - d. Analice qué variables son las más influyentes para predecir los *features* que haya seleccionado para predecir.
- 5- **Esperanza de vida.** Analice el archivo *who_life_expectancy.csv* que contiene información de diferentes países. Realice un procedimiento similar al ejercicio anterior, pero en este caso para predecir la esperanza de vida de un país (variable 'Life expectancy').
 - a. El archivo contiene muchos valores nulos. Puede eliminar todas las filas que posean valores nulos con la instrucción *dataf.dropna()*.

- b. Un modo posible de rellenar valores nulos sería creando un modelo de regresión que prediga el valor faltante en una variable. Intente realizar esta tarea eligiendo una variable para regresionar (por ejemplo '*Population*', que tiene 652 valores nulos). Para esto deberá utilizar la variable (sin los valores nulos) como *target* de la regresión. Luego utilizar el modelo generado para predecir los valores faltantes.