

**UNIVERSIDAD DE LAS FUERZAS ARMADAS ESPE**

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**



**MATERIA:**

MODELOS DISCRETOS PARA TI

**TÍTULO:**

ALGORITMOS

**AUTORES:**

Andy Reyes

David Reyes

Melany Vinueza

**DOCENTE:**

Ing. Washington Loza

Quito 3 de julio del 2023

## Índice

Introducción.....	3
Objetivo.....	3
Marco teórico.....	3
1. Algoritmos de regresión .....	3
2. Regresión lineal.....	3
3. Regresión logística.....	5
Conclusión.....	10
Bibliografía.....	10

## Introducción

Los de regresión son una herramienta importante y ampliamente utilizada en estadística y aprendizaje automático. El objetivo clave de las tareas basadas en regresión es predecir etiquetas de salida o respuestas que son valores numéricos continuos para los datos de entrada dados. Existen varios tipos de algoritmos de regresión como la regresión lineal, la regresión logística, entre otros. Estos algoritmos se utilizan para el aprendizaje supervisado y se encargan de predecir una variable dependiente (objetivo) en función de las variables independientes dadas.

## **Objetivo**

Explicar la funcionalidad de los algoritmos de regresión, los tipos de algoritmos de regresión y cómo se pueden utilizar en diferentes situaciones, en los tipos de algoritmos de regresión explicaremos con ejemplos, específicamente serán algoritmo de regresión lineal y algoritmo de regresión logística.

## **Marco teórico**

### **1. Algoritmos de regresión**

Se trata de un subcampo del aprendizaje automático supervisado que tiene el fin de crear una metodología para relacionar un cierto número de características y una variable objetivo-continua.

Algunos de los ejemplos más claros de los algoritmos de regresión son la estimación de cuánto tardará una persona en llegar a un destino, la predicción del tiempo que se mantendrá un empleado en una compañía, etc.

- Vectores de soporte regresión:

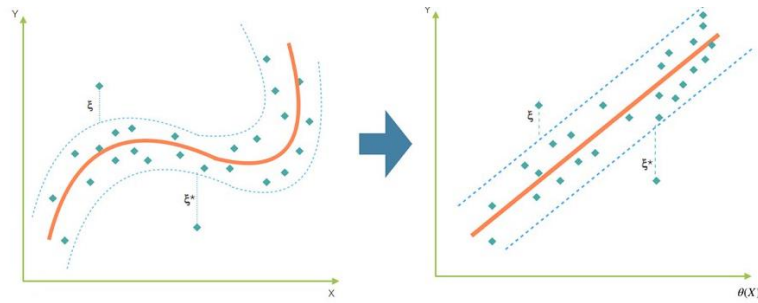


Imagen 1. Vectores de soporte de regresión. Recuperado de: <https://shre.ink/1Wpo>

Construcción de un hiperplano óptimo, de modo que el margen de separación entre las dos clases en los datos se amplía al máximo.

- Árboles de decisión regresión:

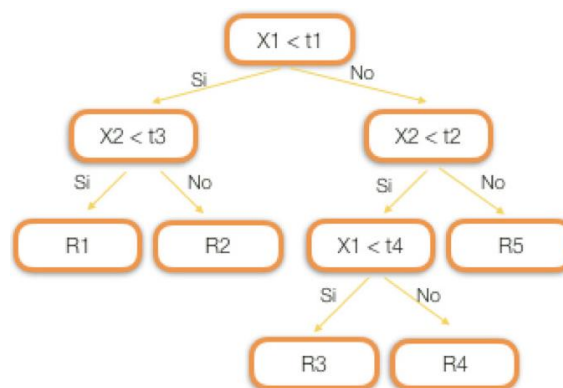


Imagen 2. Árboles de decisión regresión. Recuperado de: <https://shre.ink/1Wpo>

Dado un conjunto de datos de fabrican diagramas de construcción lógicas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva.

- Bosques Aleatorios Regresión:

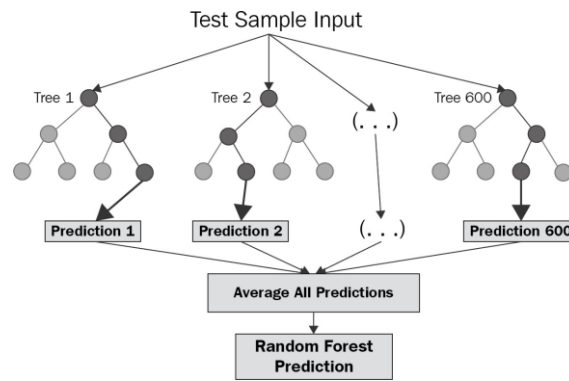


Imagen 3. Bosques aleatorios regresión. Recuperado de: <https://shre.in/IWp8>

Combinación de árboles de decisión tal que cada árbol depende de los valores de un vector aleatorios probando independientemente y con la misma distribución para cada uno de estos.

**Tabla 1.** Ventajas y desventajas Algoritmo de regresión.

	<b>Ventajas</b>	<b>Desventajas</b>	<b>Son útiles</b>
<b>Vectores de soporte de regresión</b>	<ul style="list-style-type: none"> <li>-Se pueden modelar relaciones complejas.</li> <li>-Robusto al ruido.</li> </ul>	<ul style="list-style-type: none"> <li>-Necesidad de una buena función de Kernel.</li> <li>-Requiere memoria significativa.</li> <li>-Toma demasiado tiempo para entrenar.</li> </ul>	<ul style="list-style-type: none"> <li>-Clasificación de texto o imágenes.</li> <li>-Reconocimiento de escritura a mano.</li> </ul>
<b>Árboles de decisión regresión</b>	<ul style="list-style-type: none"> <li>-Fácil de interpretar.</li> <li>-Rápido</li> <li>-Robusto al ruido.</li> <li>-Preciso.</li> </ul>	<ul style="list-style-type: none"> <li>-Difíciles de interpretar.</li> <li>-Duplicación dentro del mismo árbol.</li> </ul>	<ul style="list-style-type: none"> <li>-Diagnóstico médico.</li> <li>-Análisis de riesgo crediticio.</li> </ul>
<b>Bosques Aleatorios Regresión</b>	<ul style="list-style-type: none"> <li>-Trabaja en paralelo.</li> <li>-Maneja valores perdidos.</li> <li>-No es necesario transformar ninguna variable.</li> </ul>	<ul style="list-style-type: none"> <li>-Difícil de interpretar.</li> <li>-Más débil en la regresión al estimar valores.</li> <li>-Problemas multiclase hacia clases más frecuentes.</li> </ul>	<ul style="list-style-type: none"> <li>-Para cualquier problema de Machine Learning.</li> <li>-Bioinformática.</li> </ul>

## 2. Regresión lineal

En un sentido amplio, la regresión lineal es encontrar la relación entre las variables independientes (X) y la variable dependiente (Y). Esto significa que para ciertas variables predictoras puede obtener una relación con la variable continua que desea predecir. La

regresión lineal explica la variable Y con las variables X y obtiene la función lineal de mejor ajuste o explica esta relación.

Los términos  $\beta_0$  y  $\beta_1$  representan la intersección y la pendiente del modelo lineal, respectivamente. Estos son los coeficientes o parámetros del llamado modelo de regresión lineal. Estos dos coeficientes son dos constantes obtenidas en el llamado "entrenamiento" del modelo de regresión. Este entrenamiento se realiza utilizando datos donde conocemos los valores reales de las crestas (X,y). Una vez obtenidas las estimaciones de estos dos coeficientes, se puede predecir la variable Y utilizando la X conocida y según la siguiente fórmula:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Imagen 4. Modelo lineal. Recuperado de <https://shre.ink/IWED>

### **Mínimos cuadrados**

El método de mínimos cuadrados es un método estadístico utilizado para determinar una ecuación de regresión. Es decir, el método de mínimos cuadrados es un criterio utilizado en los modelos de regresión para minimizar el error obtenido al estimar la ecuación de regresión.

### **Ejemplo:**

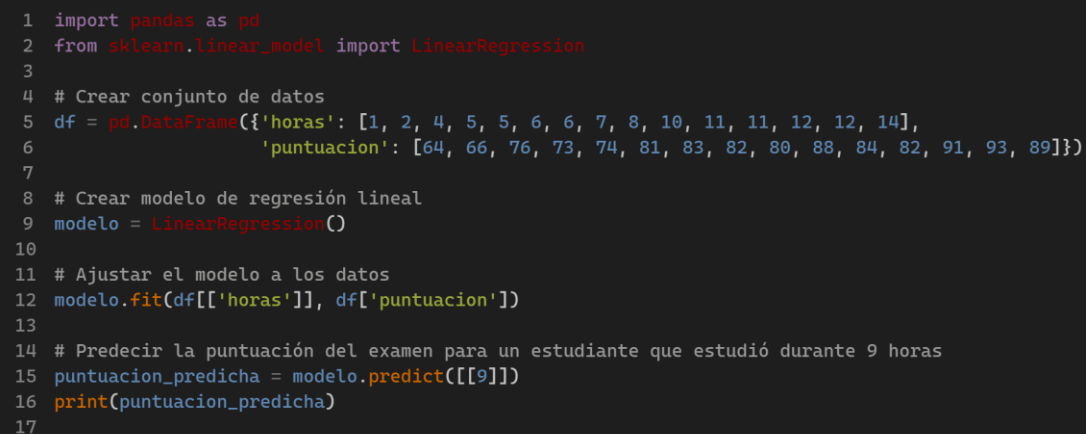
Supongamos que queremos predecir los puntajes de las pruebas en función de la cantidad de horas estudiadas. Podemos crear un conjunto de datos ficticio de 15 estudiantes con dos variables: tiempo total de estudio para algunas pruebas y puntajes de las pruebas.

Intentaremos ajustar un modelo de regresión lineal simple con las horas como variable explicativa y los resultados de las pruebas como variable de respuesta.

Algoritmo

- 1- Creamos un conjunto de datos que contiene el total de horas estudiadas y la puntuación del examen para un grupo de estudiantes.
- 2- Creamos un modelo de regresión lineal.
- 3- Ajustamos el modelo a los datos utilizando el método de mínimos cuadrados para encontrar la mejor línea recta que se ajuste a los datos.
- 4- Utilizamos el modelo para predecir la puntuación del examen para un estudiante que estudió durante 9 horas.

#### Programa Python



```
1 import pandas as pd
2 from sklearn.linear_model import LinearRegression
3
4 # Crear conjunto de datos
5 df = pd.DataFrame({'horas': [1, 2, 4, 5, 5, 6, 6, 7, 8, 10, 11, 11, 12, 12, 14],
6                     'puntuacion': [64, 66, 76, 73, 74, 81, 83, 82, 80, 88, 84, 82, 91, 93, 89]})
7
8 # Crear modelo de regresión lineal
9 modelo = LinearRegression()
10
11 # Ajustar el modelo a los datos
12 modelo.fit(df[['horas']], df['puntuacion'])
13
14 # Predecir la puntuación del examen para un estudiante que estudió durante 9 horas
15 puntuacion_predicha = modelo.predict([[9]])
16 print(puntuacion_predicha)
17
```

Imagen 5. Programa en Python regresión lineal. Recuperado de <https://shre.ink/1W0U>

### 3. Regresión logística

Es un método de aprendizaje automático el cual es supervisado y este se usa para predecir la probabilidad de que una variable ingresada pertenezca a una determinada clase-categoría, específicamente este se usa para cuando la variable únicamente puede ser de 2 clases considerándose de esta manera un algoritmo de solución binaria ya que solo puedes tener 2 clases posibles a las cuales puede pertenecer una variable.

La forma en la que funciona es que en base a una relación ya establecida el cual se obtiene por la función llamada sigmoide la cual mapea cualquier valor real en un rango entre 0 y 1, se

define como la división de 1 sobre la suma de 1 más el exponente del negativo de la entrada (valor real ingresado).

$$\frac{1}{1 + e^{-x}}$$

Ejemplo:

El ejemplo realizado demuestra la implementación básica de un algoritmo de regresión logística.



```

1 import numpy as np
2
3 # Función sigmoide
4 def sigmoid(x):
5     return 1 / (1 + np.exp(-x))
6
7 # Algoritmo de regresión logística
8 def logistic_regression(X, y, num_iterations, learning_rate):
9     num_samples, num_features = X.shape
10
11     # Inicialización de los parámetros del modelo
12     theta = np.zeros(num_features)
13
14     for i in range(num_iterations):
15         # Cálculo de las predicciones y el error
16         y_pred = sigmoid(np.dot(X, theta))
17         error = y_pred - y
18
19         # Actualización de los parámetros utilizando el descenso de gradiente
20         gradient = np.dot(X.T, error) / num_samples
21         theta -= learning_rate * gradient
22
23     return theta
24
25 # Datos de ejemplo (características X y etiquetas y)
26 X = np.array([[0.0, 1.0], [20.0, 5.5], [3.0, 50.0], [4.5, 35.0], [3.0, 22.666666666666], [2.5, 3.0]])
27 y = np.array([0, 0, 0, 1, 1, 1])
28
29 # Agregar una columna de unos para el término de sesgo
30 X = np.hstack((np.ones((X.shape[0], 1)), X))
31
32 # Dividir los datos en conjuntos de entrenamiento y prueba
33 X_train, y_train = X[:4], y[:4]
34 X_test, y_test = X[4:], y[4:]
35
36 # Entrenar el modelo de regresión logística
37 num_iterations = 1000
38 learning_rate = 0.01
39 theta = logistic_regression(X_train, y_train, num_iterations, learning_rate)
40
41 # Realizar predicciones en el conjunto de prueba
42 y_pred = sigmoid(np.dot(X_test, theta))
43 y_pred_binary = np.round(y_pred) # Redondear las predicciones a 0 o 1
44
45 # Calcular la precisión del modelo
46 accuracy = np.mean(y_pred_binary == y_test)
47 print("Precisión del modelo: {:.2f}".format(accuracy))
48

```

Imagen 6. Programa en Python regresión logística.


Al inicio se define la librería a usar (numpy) al igual que definimos 2 funciones que se utilizarán la sigmoide y la logística de regresión para los datos ingresados con sus iteraciones y el factor de aprendizaje que se tiene, dando como el “theta” el cual es la actualización de la tasa de aprendizaje y gradiente.

La línea 37 y 38 son muy importantes debido a que aquí se definen el número de iteraciones que se van a realizar para encontrar la probabilidad de lo ingresado y el “learning\_rate” o tasa de aprendizaje que es el numero el cual se va a ir multiplicando por el gradiente resultado de que se obtiene al multiplicar las matrices de los valores ingresados por el error dividiéndolo para el numero de muestras las cuales tienen la misma longitud de la matriz ingresada, mientras más baja la tasa de aprendizaje mejor será la predicción y mayor será el número de iteraciones necesarias para encontrar el valor.

## **Conclusión**

En conclusión, los algoritmos de regresión son una herramienta importante y ampliamente utilizada en estadística y aprendizaje automático. Estos algoritmos se utilizan para el aprendizaje supervisado y se encargan de predecir una variable dependiente en función de las variables independientes dadas. Existen varios tipos de algoritmos de regresión, cada uno tiene sus propias ventajas y desventajas y se pueden aplicar en diferentes campos como en precios de vivienda, predecir camas hospitalarias necesarias o identificar las cervezas que necesitarás en un inventario... etc.

## **Bibliografía**

- The Black Box Lab. (2022, June 30). *Machine Learning: Algoritmos de clasificación y regresión - The Black Box Lab*. <https://theblackboxlab.com/2022/05/06/machine-learning-diferencias-entre-algoritmos-clasificacion-regresion/>
- Gonzalez, L. (2020). Ventajas y Desventajas de los Algoritmos de Regresión.  Aprende IA. <https://aprendeia.com/ventajas-y-desventajas-de-los-algoritmos-de-regresion-machine-learning/>

*Bosque de regresión / AI Planet (formerly DPhi).* (n.d.). AI Planet (Formerly DPhi).

<https://aiplanet.com/learn/supervised-learning-regression-es/bosques-aleatorios-random-forest/1735/bosque-de-regresion>

Alvaro. (2018, 25 diciembre). *Regresión lineal en Python*. MachineLearningParaTodos.com.

[https://machinelearningparatodos.com/regresion-lineal-en-python/#:~:text=Ejemplo%20de%20regresi%C3%B3n%20lineal%20simple%20usando%20Python%201,8%20y\\_pred%20%3D%20modelo.predict%20%28x%29%20,%20M%C3%A1s%20elementos](https://machinelearningparatodos.com/regresion-lineal-en-python/#:~:text=Ejemplo%20de%20regresi%C3%B3n%20lineal%20simple%20usando%20Python%201,8%20y_pred%20%3D%20modelo.predict%20%28x%29%20,%20M%C3%A1s%20elementos)

Estadística, P. Y. (2023). Mínimos cuadrados. *Probabilidad y Estadística*.

<https://www.probabilidadyestadistica.net/minimos-cuadrados/#%C2%BFque-es-el-metodo-de-los-minimos-cuadrados>

Statologos. (2021). Cómo realizar una regresión lineal simple en Python (paso a paso).

*Statologos*. <https://statologos.com/regresion-lineal-simple-en-python/>