Gradient Descent Algorithm applied to Linear Regression

Abad L. Freddy L. , Zhagui C. Wilson I.
*freddy.abadl@ucuenca.edu.ec, israel.zhagui07@ucuenca.edu.ec

***Resumen*—Artificial Intelligence has become one of the most important branches of Computer Science today. Machine Learning, based on Supervised Learning, produces AI that resembles Human Intelligence, given input data, can produce accurate predictions. The application of AI extends to almost all areas of daily life. That is why, they must produce more robust and complex systems. To improve the AI, the bases on which it is based should be reviewed. Linear Regression is one of the bases of Supervised Learning that are proposed to study and apply in this article. Thus, an important conclusion was reached in the calculation methods of the cost function to minimize errors in Linear Regression.**

***Index Terms*—Linear Regression, IA, Machine Learning, Gradient.**

I. Introduction

Artificial Intelligence is defined as the action of a machine imitating the cognitive functions that humans associate with other human minds, such as: "learn.ªnd "solve problems."Based on elements such as Machine Learning, one of the most important branches of AI, which develops techniques that allow computers to learn automatically. (Create programs capable of generalizing behaviors from information provided in the form of examples). Classified in: Supervised Learning, Unsupervised, Reinforced, Semi Supervised, for reinforcement, Transduction, Multi Task. One of the most studied fields at present is Linear Regression, which is a mathematical model used to approximate the dependency relation between a dependent variable Y, the independent variables Xi and a random term ϵ. It is a predictive method, the most used in supervised learning. The article proposes to study the minimization of errors in the cost functions, by means of the gradient descent method, in problems applied in the linear regression.

A. *General Objective*

(I) Study the Gradient Descent Method.

B. *Specific Objective*

(I) Develop a program in MatLab, which applies the Gradient Descent Method.

(II) Clarify the operation of this method, by means of different graphs in 2 and 3 dimensions.

C. Related Works .ªnalytical, graphical and numerical methods - Numerical Methods"[8]. The article makes a study of the graphical and numerical methods of the Numerical Methods for Engineering Course, performs a complete analysis of the graphs produced. By comparison, of the current article, [8] they base their development on the Python language, totally different from the Matlab language.

II. Instance

1. *Artificial Intelligence*: Artificial intelligence (AI), also called computational intelligence, is the intelligence exhibited by machines. In computer science, an ideal ïntelligent"machine is a flexible rational agent that perceives its environment and carries out actions that maximize its chances of success in some objective or task.[5]
   Artificial Intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for the use of information), reasoning (using the rules to reach approximate or definitive conclusions) and self-correction. Particular applications of the AI include expert systems, speech recognition and artificial vision.[6]
2. *Machine Learning* Machine Learning is a scientific discipline in the field of Artificial Intelligence that creates systems that learn automatically. Learning in this context means identifying complex patterns in millions of data. The machine that really learns is an algorithm that reviews the data and is able to predict future behavior.
   Automatically, also in this context, implies that these systems are improved autonomously over time, without human intervention.[3] Because Machine Learning is a relatively new area of study, there is no definition accepted by all specialists. However, there are two very popular definitions, with which it is enough to have a clear idea of the discipline: Machine learning is the process that gives computers the ability to learn without being explicitly programmed.[4]
   The second definition is much clearer, and even gives us the general idea of how most learning algorithms work:
   Create a program that can learn to perform a task, and improve its performance in it through training (experience). [4]
3. *Simple linear regression* Simple linear regression is based on studying the changes in a variable, not random, affect a random variable, in the case of a functional relationship between both variables that can be established by a linear expression, that is, its graphic representation is a straight line That is, we are in the presence of a simple linear regression when an independent variable exerts influence on another

dependent variable. [1]
The mathematical model is given in the form $Y = aX + b$, where X and Y are variables a and b are constants (Figure 1, 2). Where "a" is the ordinate at the origin, that is, it is the height at which the line intersects the Y axis. It is also called an independent term. In addition, "b" is also called slope is the inclination of the line, that is, is the increase that occurs in the variable Y when the variable X increases one unit. [2]

$$y = A \cdot x \tag{1}$$

Figura 1. Basic Linear Regression.

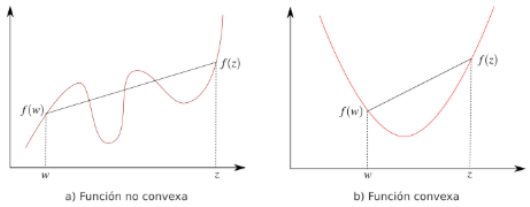

Figura 2. Cost function for non-complex cases of linear regression.

4. *Descent of the gradient* The gradient descent method in linear regression allows obtaining the adjustment curve of a function in an iterative way, minimizing the cost function [9]. Which generates combinations of each of our parameters. The problem solved by this method is the complexity involved in systems with more than 2 parameters Figure [3].
Being multidimensional, algebraically and geometrically, finding the value that minimizes the cost entails a high computational difficulty. That is why, based on the multivariate calculation, the gradient descent method applies the partial derivatives to find the variations of each dimension (axis) of the system. This method is a key component in Artificial Intelligence, specifically, in Neural Networks.
The steps followed by the gradient descent method are:

a) Randomly choose any point in the dataset.

b) Find the gradient at that point (parameter variation rate).

c) Find the highest slope, that is, the highest gradient.

d) Advance a distance 'r' in the direction of the highest chosen slope.

The formula that represents this method is:

$$\theta = \theta - \alpha \nabla f \cdot \tag{2}$$

Where:

a) θ: Parameters (Dataset).

b) α: Learning Ratio (Defines how much the gradient affects, the update of parameters in each iteration).

c) ∇f: Gradient.

This process is repeated until finding a ∇f similar to 0, which represents a minimum point or point of inflection of the system. To find all the local or global minim of the system, the process is repeated a certain number of times.
The learning ratio is usually stipulated as input. A problem that generates, delegate to the learning ratio as an input, is the bad approach that generates.
It is a fact that generates large differences, since being too large a value (range of 0-1), produces large advances, which are less reliable approximations in the minimum. And if the learning ratio is very small, progress to the minimum is more reliable but much slower. Some techniques used to avoid this problem are:

I SGD.
II Momentum.
III NAG.
IV Adagran.
V Delta wing.
VI Rmsprop.

In developed practice, a learning ratio is taken as input data, based on an average use value.

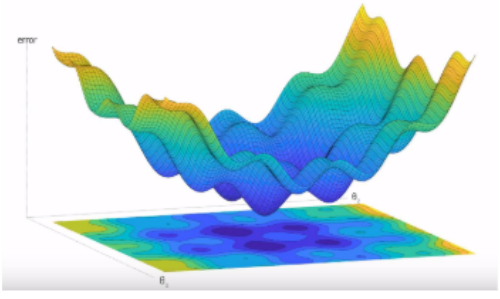

Figura 3. Data modeling with variation of parameters and complex derivation.

5. *Level curves* The system of representation of contour lines consists in cutting the surface of the land by means of a set of planes parallel to each other, separated a certain distance from each other. Each plane cuts to the ground forming a figure (flat) that is called level curve or isohipsa. The projection of all these level curves on a common plane (the map) gives rise to the desired representation.[7]
   Level curves, in developed practice, allow a clearer understanding, being projections of multidimensional systems, geometrically impossible to graph or difficult to understand. Location of Points where Gradient tends to 0 (Figure 4).
   Level curves show the minima of the system as densely populated points, by the minimum variations between each approach, closer and closer (Figure 5).

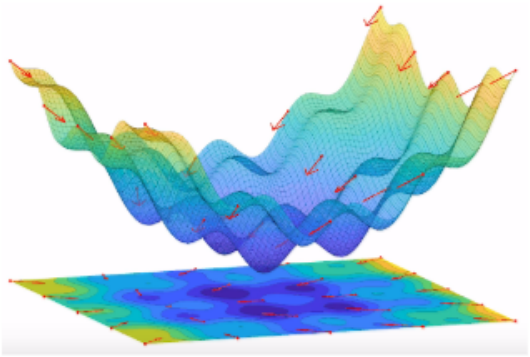

Figura 4. Location of Points where Gradient tends to 0.

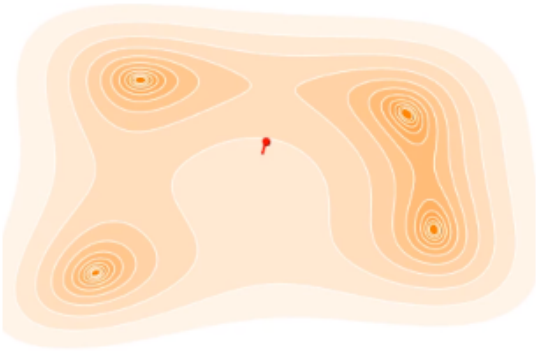

Figura 5. Level Curves, with the minimums of the cost function.

III. Methodology

Tado he gradient descent method was applied in the Matlab programming language, using the Octave development IDE, for its advantages of dynamic memory management. The input data, exemplify the relationship between the average population of a city, with the average monthly income of each inhabitant. This figure is scaled to 10^{-5} of the figure. And it allows, for example, to find the per capita income of the city. The development of the code in Matlab, is summarized in:

I Reading of the input data (theta), and the input ratio.

II Assignment to a processing function, the input data.

III Application of the formula 2.

IV Graph of dataset points and linear regression with minimum cost function.

V Prediction of data, based on this processed data.

IV. Results

The results obtained in this work allow us to verify the processing advantage of this method. Being thus Figure [7], the input data. Figure [10] modeling of the input data with a random function. Figure [8] and [9] graphs of input data and linear regression function with minimum cost, in 2 Dimensions and 3 Dimensions, respectively. Concluding thus, in a fairly close prediction before an entry data. Figure [10]

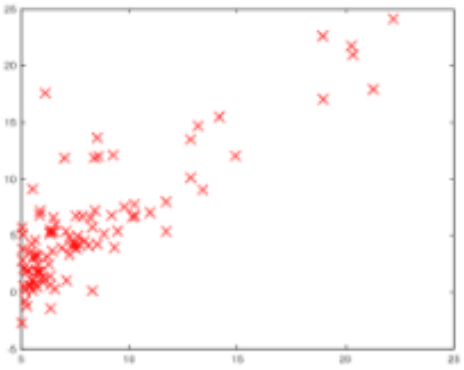

Figura 6. Input data chart.

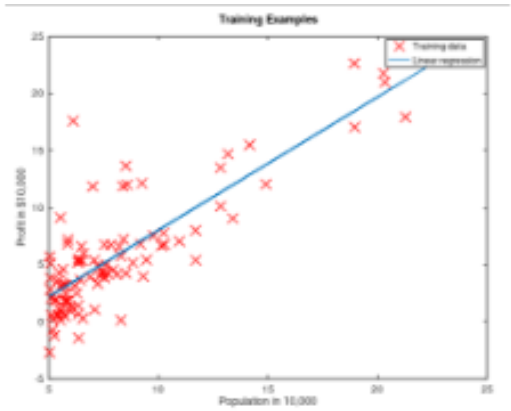

Figura 7. Data graph and linear regression in 2 dimensions

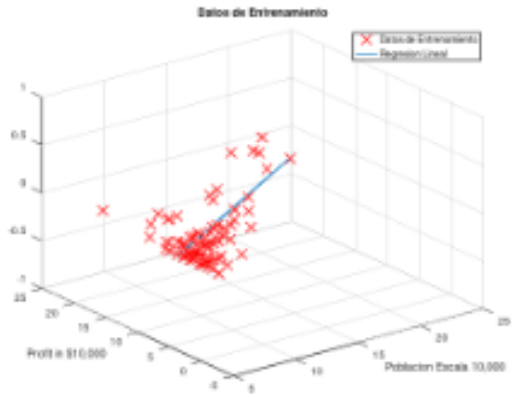

Figura 8. Level Data graph and linear regression in 3 dimensions

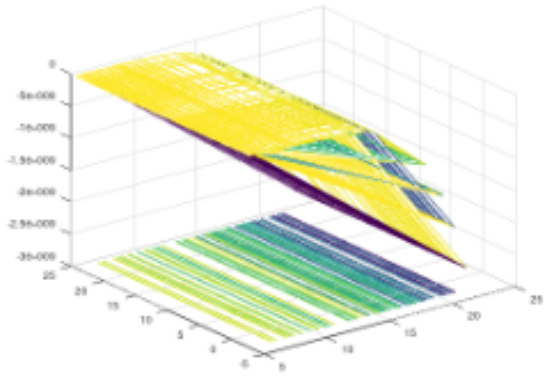

Figura 9. Modeling of the input data, under a function.

```
Alpha : 32.0727

Prediccion para 35000
Prediccion para 3.5:    2.246763

Prediccion para 70000
PPrediccion para 7.0:   5.355739
```

Figura 10. Prediction results of input data for practice.

V. Conclusions

1. The gradient descent method performs an optimal processing compared to the common method of least mean square error.
2. The learning ratio, is an element that must be carefully monitored, the choice of this, causes an optimal development of the gradient descent method.
3. Linear regression is the elementary basis of Machine Learning Supervised Learning, perfecting the techniques for its development guarantees complete AIs.

Referencias

[1] Anonimo, "*Regresion Lineal*", Anonimo, 2016
[2] Anonimo "*Regresion Lineal*", Piramide, Madrid, 2017
[3] Gonzales Andres, K.D. Fink, "*Machine Learning*", Prentice-Hal, 2016
[4] Pasillas Adric, "*Machine Learnin- Guia para Principiantes*", Pasillas, 2017
[5] Editores Wikipedia, "*Inteligencia Artificial*", Wikipedia, 2018
[6] Rousse Margaret, "*Inteligencia Artificial*", Rousse, 2016
[7] Urrutia Javier, J.M.REY, "*Curso de Cartografía, Orientación y GPS*", Piramide, Madrid, 2011
[8] UVG-mm, K.D. Fink, "*Métodos analíticos, gráficos y numéricos - Métodos Numéricos*", Prentice-Hal, 2010
[9] Sanchez Alejando, "*Algoritmo de descenso de gradiente*", Wiki, 2018