Gradient Descent for Prediction of Crop Yield

Abstract

This report aims to explore the application of gradient descent to determine the optimal fertilizer amount for increased crop yield. Microsoft Excel was used to implement the gradient descent algorithm in order to understand the functionality behind each step toward minimizing the loss function. The study of the fertilizer amount added to crops and the resulting crop yield indicates that to a certain extent, crop yield increases with fertilizer amount. The findings emphasize the role that gradient descent can play to provide data-driven advice to farmers as the minimized loss function will account for accurate crop yield prediction.

Introduction:

Gradient descent is an iterative algorithm that is used to find and optimize the values of the parameters of a function which can be used to minimize the cost function of a model. An error value is generated after a prediction as there is a possibility for a difference between the actual output and the output predicted by the model. A cost function of a model or algorithm indicates the average of the total errors of all the training samples.

The dataset that was utilized for performing gradient descent using Microsoft Excel included the amount of fertilizer in kilograms per hectare provided to crops and the crop yield in kilograms per hectare. Fertilizers are generally added to improve crop growth and productivity, however, the relationship between the independent fertilizer amount and the dependent crop yield is not always linear. An inadequate amount of fertilizer supply can reduce crop yield and also cause stunted growth in crops while an excessive amount can also lead to reduced crop yield due to nutrient imbalance and environmental pollution. The studied dataset contained values of inadequate and nearly optimal amounts of fertilizer which leads to reduced crop yield and an improvement in crop yield, respectively. The dataset did not comprise information regarding the presumed drop in crop yield that will occur as a result of over-fertilization. The dataset was generated by a chatbot, ChatGPT, based on specified criteria regarding the established link between fertilizer amount and crop yield.

Methodology:

The procedure to be adapted for using gradient descent to optimize the model’s performance is as follows:

1. Firstly, arrange the dependent and independent variables and ensure that there is a correlation between these variables.
2. To perform linear regression, first the hypothesis function is defined and the one employed for this task was H(x) = W1x1 + W2x2 + b.
3. Then using the formulae x1 = x^2\*s1, x2 = s2\*x, and y0 = sy\*y, where s1, s2, and sy are constants used for scaling the x1, x2, and y values, the values of x1, x2 and y0 are calculated which are used in the program for further crucial calculations.
4. Next, the loss function is defined as L (x, y) =1/2(H(x) – y0) ^2. This is calculated to account for how good the chosen hypothesis function is by checking the difference between the predicted and actual results and then reducing it.
5. Following this, in order to tune the parameters to optimize the gradient descent, the derivative of the loss function is calculated using the formulas,

dL/dW1 = (H(x) – y0) x1

dL/dW2 = (H(x) – y0) x2

dL/db = H(x) – y0

1. W1, W2 and b are taken as particular constant values for the first step. Then using the formulas,

W1.new = W1 – a(avg(dL/dW1))

W2.new = W2 – a(avg(dL/dW2))

b.new = b – a(avg(dL/db)),

we update the parameters in the right direction where ‘a’ is the learning rate of the manually created model. ‘a’ is a constant value while the averages of the derivatives of W1, W2, and b are calculated in order to tune the parameters.

1. These mentioned steps are repeated for each newly tuned set of W1, W2, and b values. In the end, the main objective is to ensure that the loss function is minimized.

Discussion:

Fig. 1.1

Fig 1.2

Fig 1.1 shows the original graph of the fertilizer amount on the x-axis plotted against the crop yield on the y-axis while Fig 1.2 shows the curve of the minimized loss function which has been brought close to the original plot of the two variables in order to demonstrate the function of the gradient descent algorithm. The loss function is observed to reduce and eventually reach a minimum value. After the graph for the 20th step, the curve is noted to take up a constant and minimized figure.

Conclusion:

In conclusion, the study conducted on fertilizer amount and crop yield revealed that there is a positive correlation between the variables. The reduced loss function of the gradient descent algorithm confirms that the optimal amount of fertilizer required for maximum crop yield can be predicted using the created linear regression model in Microsoft Excel.

References:

[1] P. Pandey, “Understanding the Mathematics behind Gradient Descent.,” Medium, Jan. 17, 2020. <https://towardsdatascience.com/understanding-the-mathematics-behind-gradient-descent-dde5dc9be06e>

[2] “Google Drive: Sign-in,” accounts.google.com. https://drive.google.com/file/d/1ABQKdd8oTJLnCdqDB5MZhV-ndQcW1kNN/view (Accessed May 02, 2023).

[3] “Generate a real-life dataset of fertilizer amount and crop yield with 50 entries in a table format without programming”, ChatGPT Mar 23 Version, OpenAI, Apr. 27, 2023, <https://chat.openai.com/>