MODELLING OF POLYNOMIAL REGRESSION ALGORITHM

in python using Verilog Modules

Abstract:

Aim: This project is a software implementation of Polynomial Algorithm in Python by using modules of verilog.

Methodology: Algorithm has been written in python. So, it is the heart of this project. It is basically providing some library from which implementation of algorithm has become easy and we can also visualize it with the help of Graph of the results. All the computation such as additions, multiplication etc. are being done using verilog modules whose test bench is generated using python and its output is being taken using System calls

Practical Implementation: With this project, we are showing how to implement AI algorithm in hardware to bring efficient performance.

Findings: we are training the machines to find the output value for an input value according to the given data-sets which is firstly it find the pattern in the data-sets and give the results accordingly.

CONTENTS

■ Introduction

Polynomial regression is a form of regression analysis i.e predictive modelling in which the relationship between the independent variable X and the dependent variable Y is modelled as n^{TH} degree polynomial in X.

It's almost the same as linear regression but for polynomial data set, unlike linear regression, it fits a nonlinear relationship between the value of X and the corresponding conditional mean of Y,denoted E(Y/X). Although polynomial regression fits a nonlinear model to the data, as the statistical estimation problem it is linear, in the sense that the regression function E(Y/X) is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression.

■ Its representation

In general , we can model the expected value of y as nTH degree polynomial, yielding the general polynomial regression model
$Y = B_0 + B_1X + B_2X^2 + ... + B_nX^n + E.$
Conveniently, these models are all linear from the point of view of estimation, since the regression function is linear in terms of the unknown parameters $B_0,B_1,.....$ Therefore, for least squares analysis, the computational and inferential problems of polynomial regression can be completely addressed using the technique of multiple regression. This is done by treating $X,X^2,...$ as being distinct independent variables in a multiple regression model.

■ Matrix form and calculation of estimates

The polynomial model

$Y_i = B_0 + B_1X_i + B_2X_i^2 + ... + B_nX_i^n + E.$ (i = 1,2,3,4,...n)

Can be expressed in matrix from in terms of a design matrix X, a response vector Y, a parameter vector B, and a vector e of random errors. The i-th row of X and y will contain the x and y value for the i-th data sample. Then the model can be written as a system of linear equations:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^m \\ 1 & x_2 & x_2^2 & \dots & x_2^m \\ 1 & x_3 & x_3^2 & \dots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^m \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Which mean using pure matrix notation is written as

Y = XB + E

The vector of estimated polynomial regression coefficients (using ordinary least squares estimation) is

$B = (X^T X)^{-1} X^T Y$

■ why we use polynomial regression

In real world, outputs may not be directly proportional to inputs such as in the case of buying things such as chocolates.

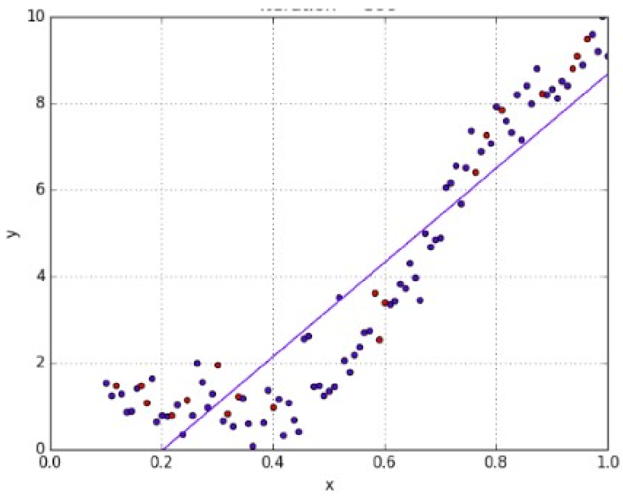
Suppose one chocolate costs Rs 5, then 20 chocolates will cost Rs 100. This is not the case that such a linear relationship holds in every situation. For example, if we are modeling the rate of increase/decrease of population in some country then we may find that each year the rate of increase/decrease in population may not be linear i.e suppose the population of a country was increasing and then due to some natural disaster many people of that country died. then in this case if we plot the graph of population verses time, we can see that the graph is not linear hence in this case, we can use polynomial regression to roughly estimate the population. In this case, we might propose a quadratic model of the form $Y = \theta_o + \theta_1 X_1 + \theta_2 X_1^2$.

Some more applications of polynomial regression are:-

❏ Growth rate of tissue

❏ Progression of disease epidemics

❏ Distribution of carbon isotopes



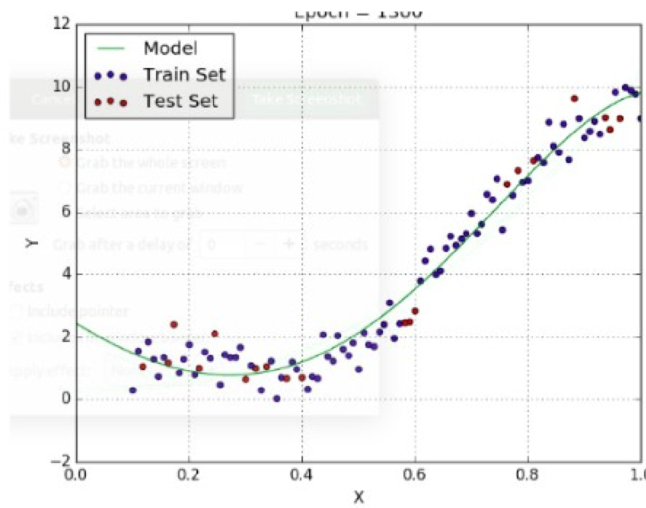


Fig:1) Graph of Linear Regression   Fig:2) Graph of Polynomial regression

Here we can clearly see that polynomial regression is showing the graph which is more closer to the data sets while linear regression is somewhat deviated from the data sets. But if we had divided the data sets into some parts and then we find the estimated value of data sets then it will be closer to the polynomial regression that's why, sometimes it is also called special case of multiple linear regression.

■ Algorithm working

For finding estimate value in polynomial regression,

- We find the mean of x and y, in which we are using recursive doubling N no of times, where N is the size of data.
- Then we are calculating means of X^2 , X^3 and X^4, for these we are using multiplier and recursive doubling, N number of times for each.
- We are calculating product of X and Y, which is given to us and summing all these and dividing by N is giving the mean of XY, this is being calculated using MAC.
- We are calculating product of X^2 and Y, which is given to us and summing all these and dividing by N is giving the mean of X^2Y, this is being calculated using MAC as well as multiplier for squaring X
- Now we are subtracting square of mean of X from mean of X^2.
- Then we are subtracting multiplication of means of X and Y from mean of XY.
- Then we are subtracting multiplication of means of X^2 and X from mean of X^3.

- Then we are subtracting multiplication of means of X^2 and X^2 from mean of X^4.
- Then we are subtracting multiplication of means of X^2 and Y from mean of X2Y.
- We are doing the above 5 steps, to calculate the coefficients in the polynomial.
- Using all these, we are predicting the value of Y.

- For calculating the Mean,we are using N numbers of Recursive doubling Adders.
- For finding the mean of Products,we are using 16 bit wallace multiplier.
- We are using MAC for accumulating values of product of X and Y, The module used for MAC is 16bit mac.
- We are using Subtraction module for calculation of coefficients of polynomials.
- Finally we are showing the graph of actual output and predicted output

■ Modules used

N is 12 , according to our data Set

- ➔ Number of Full Adder(64 bit): 2*N
- ➔ Number of Recursive doubling(64 bit): 6*N + 3
- ➔ Number of MAC(16 bit): 2
- ➔ Number of Subtraction(64 bit): 10
- ➔ Number of Wallace Multiplier(16 bit): 10 * N + 18

■ Contributors

This documentation has been contributed by two students of **IIITDM Kancheepuram** under the guidance of Our teacher **Dr. Noor Mahammad**. First student is **Amar Kumar**(Roll no. - CED17I029) and Second student is **Shiv Shankar**(Roll no. - CED17I026). Both are pursuing Computer Engineering(2017-2022).