Estimating Economic Growth using Deep Learning

First Author Name[a*], Second Author Name[bc], etc.[ac]
[a]Department, University, City, Country
[b]Department, University, City, Country
[c]Department, University, City, Country

[*]Corresponding author: first name, initials, surname; email.address

Abstract

The economic growth rate of countries is usually computed in terms of GDP per capita, which is largely influenced by macroeconomic factors, as depicted by certain case studies and papers. The aim of our paper is to present a model that predicts the economic growth as the GDP per capita of a country. We will be using data from the Data Bank for World Development Indicators of the World Bank. Post this, we will use python for data analysis, and then build a deep learning model using Artificial Neural Networks on the TensorFlow framework using the Keras library. We will also try to reason the findings by the correlation matrix and scatter plots and compare them with the previously done research and find similar patterns.

Keywords: keyword 1; keyword 2; lower case except names, max 6

1 Introduction

Economic growth can be defined as the increase in the inflation-adjusted market value of the goods and services produced by an economy over time. It is generally measured as the percent rate of increase in real gross domestic product, or real GDP. The economic growth rates of countries are usually computed in terms of the percentage increase in GDP per capita. Conventionally, our standard of living is measured by the quantity of goods and services accessible, thus we may even say that economic growth leads to an increase in our general standard of living.

However, research into the field of economic growth tells us that growth cannot occur in isolation. There exists a dependency of growth to a multitude of variables. It is a widely accepted notion between scholars that GDP of a country acts as a good indicator of economic growth; the higher the GDP of an economy, the more vigorous its growth. There exist many macroeconomic factors that directly influence the growth rate of a country's GDP as well.

Shiva S Makki's and Agapi Somwaru's published works in the American Journal of Agricultural Economics states how factors like foreign direct investment, trade, labour force, etc, directly influence the economic growth in developing countries. We roughly see a similar relation apply to other countries with different development levels as well.

On the basis of these hypotheses, we aim at building a model that estimates the economic growth, as changes in GDP per capita, of a country based on certain relevant factors, using artificial neural networks.

2 Literature Review

Many publications suggest a correlation between change in GDP per capita of a country with a myriad of macroeconomic factors. Such macroeconomic evidence not only provides dominant support for how macroeconomic factors have a positive influence on a country's economic growth, but also tells us the significance these factors hold in determining the state of development of a country. For instance, there is a correlation between economic growth and rising international trade; it has been found that in the last

200 years, the world economy has experienced sustained economic growth accompanied by even faster growth in global trade.
Certain studies have also shown an increase in Foreign Direct Investment results in higher economic growth of a country. Evidence from both developing and developed countries tell us how FDI, not only promotes economic growth, but also indirectly influences economic growth via certain interaction terms, that interacting with human capital exerts a strong positive effect on economic growth.
Similarly, different case studies concerning particular countries from all over the world, also show a negative correlation between unemployment rate and change in GDP per capita of a country - which aligns well with the hypothesis suggested in Okun's Law. Similar case studies can also be found for different countries, when it comes to other factors like human capital index, expenditure of education, etc, and their influence of the economic growth of a country.
These set of variables have been chosen from different sectors like Education, Finance, Health, Infrastructure, Poverty, Environment. These factors have also been chosen on the basis of their influential capacity over GDP per capita, as provided by a collection of different approved studies - as mentioned in the references section. All of these factors have different intensities of significance when it comes to their influential power over altering the GDP per capita of a country.
In this paper, we attempt to illustrate, via the means of a DL model, how such a varied set of factors can predict economic growth quantitatively, as the change in GDP per capita of a country.

3 Objectives

The objective of this research is to use artificial neural networks to predict the economic growth (measured as changes in GDP per capita) of a country, with reasonable accuracy, considering the data of 264 countries over 20 years. Our model will encompass 20+ factors including gross savings, unemployment rate, education expenditure, profit tax, the amount of foreign direct investment and other factors mentioned in the table of description of variables. Our main motive will be to train the deep learning model such that it predicts the change in GDP per capita with least possible mean absolute error and analyse the dependence of variables and reason them using previously done research.

4 Hypothesis

Studies from several reputed journals depict a correlation between macroeconomic factors and economic growth. The premise of our hypothesis is to bring forth the idea that the real GDP, or the changes in GDP of a country, does not depend on the name of the country or the year, but on the actual variables which directly or indirectly affect the GDP. These variables or factors have been chosen from various domains like health, literacy, infrastructure, etc, that are widely known to potentially affect the GDP.
In our study, we worked around two sub hypotheses in order to ascertain the target variable of our model, and to test and verify the central hypothesis of this paper. We conjecture that, our model, that disregards both the year and the name of the country, would not predict "changes in GDP" as well as it would predict only "the GDP", given we take into consideration exactly the same variables. We make this conjecture based on the fact that the variables we took into consideration did not give us the expected changes - as we only used data for a particular year, and not the respective net changes for each of the variables. We'll build models that predict both changes in the GDP and the GDP alone, and note their observations to test the credibility of our conjecture.
To sum up, we hypothesize that the GDP can be predicted more efficiently than the changes in GDP by our model, without using the name of the country or the year. Furthermore, we will also try to see the dependency between each of these features with the help of a correlation matrix and scatter plots.

5 Table of description of variables

Variable	Interpretation	Source
GDP per capita	The gross domestic product per capita, or GDP per capita, is a measure of a country's economic output that accounts for its number of people. It divides the country's gross domestic product by its total population.	World Bank
The amount of foreign direct investment	A foreign direct investment (FDI) is an investment made by a firm or individual in one country into business interests located in another country. FDI measures the total level of direct investment at a given point in time, usually the end of a quarter or of a year.	World Bank
Gross Savings	Gross saving is disposable income less consumption. It can be calculated as gross national income less total consumption, plus net transfers.	World Bank
Unemployment Level	% of people from the labour force who are unemployed.	World Bank
Labour force	% The labor force is the number of people who are employed plus the unemployed who are looking for work.	World Bank
Child Mortality Rate (per 1000 live births)	% The probability of a child born in a specific year, dying before reaching the age of 5.	World Bank
New Business Registered	% Business registration certificate issued by the Department of the Treasury or such other form or verification that a contractor or subcontractor is registered with the Department of Treasury.	World Bank
Education Expenditure (USD)	% Represents a part of the government's budget allotted to different educational activities, for any country.	World Bank
Age Dependency Ratio	% The ratio of the number of dependents aged 0-14 and over 65, to the total population aged 15 to 64.	World Bank

6 Methodology and methods

We will be collecting the data from the Data Bank for World Development Indicators of the World Bank. We will use the features of the past 20 years of 264 countries. We will intentionally not be using the year and the name of the country in our feature set, as we want to measure the change in GDP per capita irrespective of the year and the country, and just on the basis of the other variables. We will be using python to do all of the data analysis and implementing the deep learning model.

After overviewing the data, we found some missing values in the data. We will fill these values with the average of the non-NaN values of the particular feature. After the NaN values have been removed, we will be analysing the relationships between the variables using the correlation matrix. We will try to reason the findings by the correlation matrix and compare them with the previously done research and find similar patterns. We will also visualise the data using scatter plots to do the same. We will also extract some new variables from the available variables like the change in GDP per capita from the GDP per capita of the countries.

The dataset will be divided into the feature set (X) and the target variable (y). The feature set can be seen from the table of description of variables above and the target variable will be the change in GDP per capita which is an indicator of economic growth. We will be label encoding and one-hot encoding the categorical variables so that our model can use these features effectively. We will normalise the non-categorical variables using the Standard Scaler to make the mean and standard deviation of each feature 0 and 1 respectively. We will also try to do Principal Component Analysis and reduce the dimensionality of the feature set by grouping similar features together. Let the final number of features after doing PCA be n. We will split our data into train, validation and test sets in the ratio 80:10:10. We will use the training data to train our model, the validation data to tune the hyperparameters of the model, and the test data to measure the mean absolute error of our model.

Finally, we will be building a deep learning model using Artificial Neural Networks on the TensorFlow framework using the Keras library. The following network architecture and the hyperparameters are tentative and subject to changes based on the performance of the model on the validation dataset. We plan to build the model with n input units, 2 hidden layers having 256 units each and an output layer having 1 unit which gives the estimated change in GDP per capita. We will use the Rectified Linear Unit (ReLU) activation function for each unit. We will be using stochastic gradient descent using the Adam optimizer and use the mean absolute error as the loss function as well as the metric for our model. We will also be using dropout and batch normalization to reduce overfitting. We will be training the model for about 50 epochs with early stopping in order to prevent overfitting. We will choose a batch size of about 256 for the gradient descent.

Link to the code: https://github.com/yashbg/EstimatingEconomicGrowth

7 Mathematical Analysis

7.1 Correlation

A correlation is a statistical measure of the relationship between two variables. The measure is best used in variables that demonstrate a linear relationship between each other. The correlation coefficient is a value that indicates the strength of the relationship between variables. The coefficient can take any values from -1 to 1. The interpretations of the values are:

- -1 : Perfect negative correlation. The variables tend to move in opposite directions (i.e., when one variable increases, the other variable decreases.
- 0 : No correlation. The variables do not have a relationship with each other.
- 1 : Perfect positive correlation. The variables tend to move in the same direction (i.e., when one variable increases, the other variable also increases.

$$r_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}}$$

where,
r_{xy} – the correlation coefficient of the linear relationship between the variables x and y
x_i – the values of the x-variable in a sample
$\overline{x}$ – the mean of the values of the x-variable
y_i – the values of the y-variable in a sample
$\overline{y}$ – the mean of the values of the y-variable

7.2 Forward Propagation

$$a^{(l)} = \sigma(W * a^{(l-1)} + b)$$

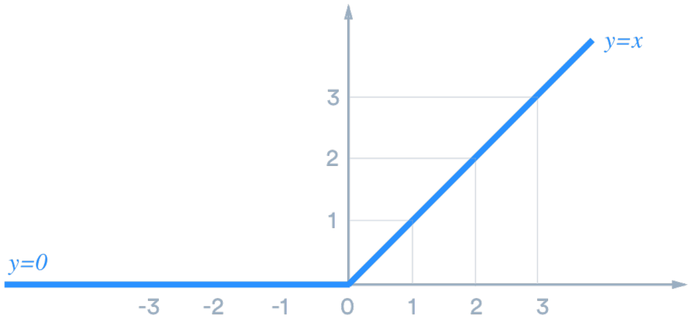
where,
$a^{(l-1)}$: activation values of the previous layer. Dimension: [number of nodes in layer (l-1), 1]
σ : the activation function being used (relu, sigmoid etc)
$a^{(l)}$: activation values of the current layer. Dimension: [number of nodes in layer (l), 1]
b : bias values of current layer. Dimension: [number of nodes in layer (l), 1]
W : weights from previous layer to the current layer. Dimension: [number of nodes in layer (l), number of nodes in layer(l-1)]
The information provided by the input layer propagates forward through the hidden layers, and produces the output. The architecture of a neural network consists of depth, width and activation function used. Depth is the number of hidden layers, while width is the neurons in each hidden layer.

7.3 ReLU activation function

The rectified linear activation function or ReLU is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses ReLU is easier to train and often achieves better performance. The ReLU is the most used activation function in the right now. The function and its derivative both are monotonic.
Formula:

$$f(x) = max(0, x)$$

Graph:



7.4 Batch Normalization

$$\hat{x} = \frac{x - E[x]}{\sqrt{Var[x]}}$$

where, x = activation values from previous layer with dimension : (number of nodes in the layer, 1)
Weights in deep neural networks are very sensitive towards any minor changes. Thus, if there is a high contrast in activations, weights are going to change rapidly and may cause overfitting.That is why batch normalization is a very important technique used in training deep neural networks. As we receive the activations from the previous layer, before going for a forward propagation, we normalize the whole layer by making the mean of the layer = 0, standard deviation = 1. This helps standardize the input before going to the next layer. This helps in training the network faster, makes it easier for training deep neural networks.

7.5 Mean Absolute Error

The Mean Absolute Error (MAE) is the average of all absolute errors. The formula is:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \mid x_i - x \mid$$

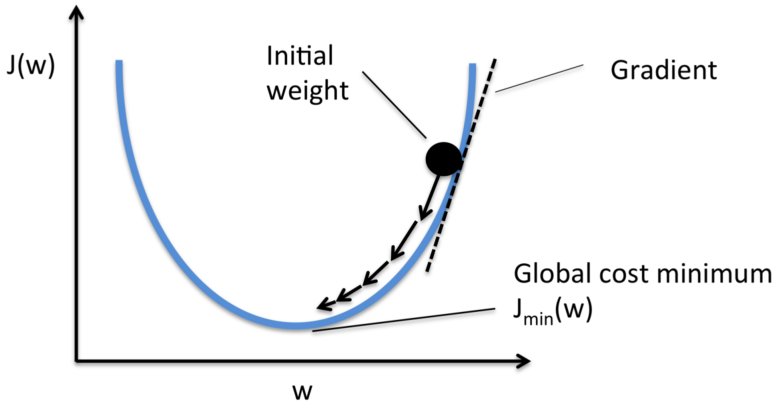
Where:
n = the number of errors,
$\sum$ = summation symbol (which means "add them all up"),
$\mid x_i - x \mid$ = the absolute errors.

7.6 Gradient Descent

The concept for gradient descent is very basic. We have a function J(w) which depends on weight w. To get the lowest possible value of J, we first find its gradient, and move along the gradient. We can move in 2 directions, but for our purposes, we will move in the direction where J decreases. The amount of shift we take is dependent on a hyperparameter called learning rate, α. The learning rate should be tuned with the model, as too low learning rate will take a lot of time to train, and too high learning rate can overshoot the value of J and will never lead to its minimization.

$$w_i = w_i - \alpha \frac{\partial J}{\partial w_i}$$

$$b = b - \alpha \frac{\partial J}{\partial b}$$



We will be using stochastic gradient descent which takes a small sample of the training data, called a minibatch, and optimizes the hyperparameters using gradient descent.

7.7 Back Propagation

Back-propagation is the essence of neural net training. It is the method of fine-tuning the weights of a neural net based on the error rate obtained in the previous epoch (i.e., iteration). Proper tuning of the weights allows you to reduce error rates and to make the model reliable by increasing its generalization. Backpropagation is a short form for "backward propagation of errors." It is a standard method of training artificial neural networks. This method helps to calculate the gradient of a loss function with respect to all the weights in the network.

$$\delta^L = \Delta_a C \cdot \sigma'(z^L)$$
$$\delta^l = (w^{l+1})^T \delta^{l+1} \cdot \sigma'(z^l)$$
$$\frac{\partial C}{\partial b_j^l} = \delta_j^l$$
$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-1} \delta_j^l$$

where,
z^L= weighted input
a^l = activations in layer l
C = cost function
w^l = weights in layer l
b^l = biases in layer l

7.8 GDP Deflator

The Gross Domestic Product (GDP) estimates the changes in prices of all the goods and services that may be produced in an economy. It is a measure of inflation which is widely accepted and utilised by economists, as it helps in comparing the inflation adjusted economic activities or variables from one year to the other. Here, the effect of price changes on GDP is shown by, first, establishing a base year and, secondly, by comparing current prices to prices in the base year.

$$GDP\ Deflator = \frac{Nominal\ GDP}{Real\ GDP} \cdot 100$$

Where, Nominal GDP = the value of goods and services unadjusted for inflation Real GDP = the value of goods and services adjusted for inflation

7.9 K-Nearest Neighbours (KNN)

The KNN is a Machine Learning algorithm that can be used for classification and regression predictive problems. KNN helps classify a data point based on how its neighbours are classified. In KNN, the K refers to the number of nearest neighbours that one must include in order to classify a data point. In our model, we use the KNN imputer to impute the NaN values in our data.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

In this equation for the Euclidean distance, Xi and Yi refer to the coordinates of the data point.

7.10 Coefficient of Determination (R Squared)

R-squared helps us determine how well our regression line fits out data, with a value that lies between 0 and 1. The given R Squared formula compares the fitted regression to the possible 'worst' case model.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}}$$

Where,
RES = Sum of squares of residual

$$SS_{RES} = \sum_{i=1}^{n}(y_i - f(x_i))^2$$

TOT = Total sum of squares

$$SS_{TOT} = \sum_{i=1}^{n}(y_i - \overline{y})^2$$

8 Results

9 Conclusions

References