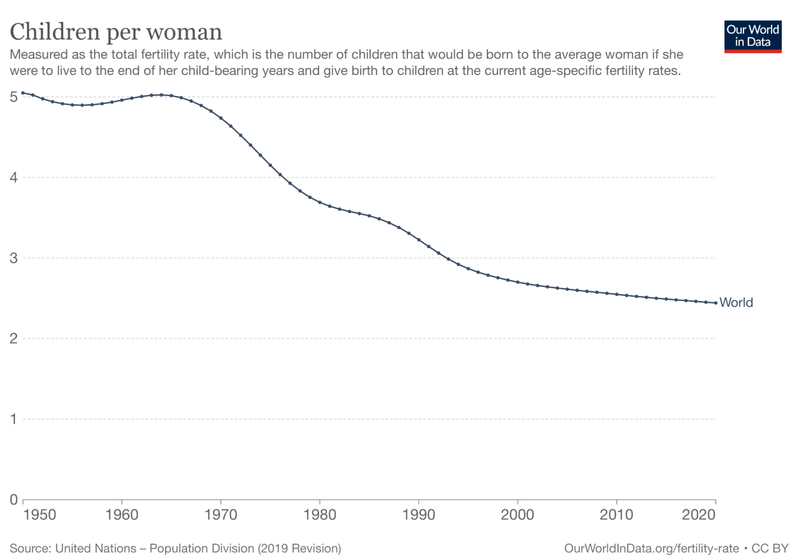
Mitchell Morrison
Gabriel Mongaras
Christian Gould
Braiden Hook

Dr. Fontenot - CS-5394

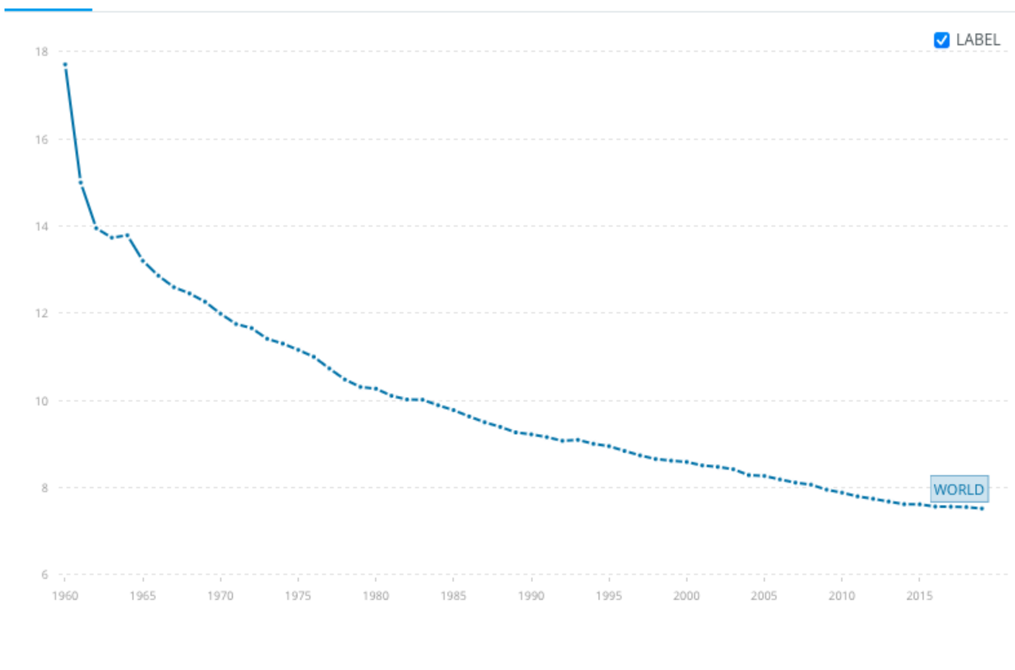
Project 1 Analysis - Human Population

Initially when this project was assigned, the task seemed simple enough - predict the population of humanity on Earth 100 years from today. However, after doing some initial research, it became evident that the current human growth rate of 1-2% annually over the last century is not sustainable given our current technological state and our finite resources on our planet. An annual growth rate of roughly ~2% reflects an exponential curve - which over the course of the next 100 years our current 7.9 Billion people would grow to be near 57 Billion humans on planet Earth should this trend continue.

Obviously there is no way that we will be able to sustain that many people on Earth under current conditions. There are many factors that need to be taken into account other than simply the past century's population data. We need to take another approach if we want to make a prediction that will have any sort of validity. If we double at the same rate we have seen in recent history (we doubled between 1950 and 1987), this would mean that in the years 2022 and 2059 we would have 7.8 and 15.6 Billion people on Earth respectively. But simply looking at the numbers of population alone does not allow us to properly see the whole picture. Fertility rate has gone down and the death rate has declined. These are three factors which have a major impact on how our world population is changing, which are easy to miss. From the graph to the left, it is clear that the fertility rate is declining more and more as the years pass by. [1]

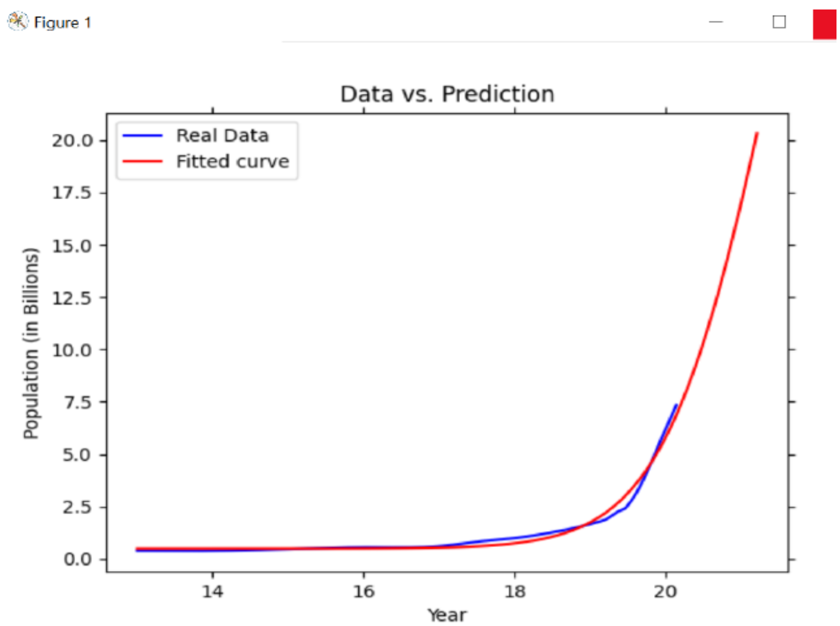


But, why then is the population of the world not decreasing rapidly? In fact, why has the population of the world not decreased, and instead has continued to increase over the course of the past few decades? The answer to that comes from looking at the death rate. The graphic below depicts the death rate per 1000 people over the course of about the last century. [2]



So, with the decreasing fertility rate, the main reason we are still increasing in global population is that our death rate is rapidly decreasing. However, as you can clearly see from the graph, our death rate is reaching an asymptote. This means that in the near future, we will meet a point at which our global population will be making people just as quickly as we are losing them. This assumption is the basis for how we chose the shape of our model - the sigmoid curve.

Our algorithm was initially implemented using a neural network which did not work super well. The neural network was trained on a dataset that contained pretty much only a straight line, so the network learned to make a HUGE straight line, and the result was not near what we wanted. The prediction the neural network gave would just grow in magnitude the more times we let it iterate. So, we turned to other ways of approaching the problem. While some members were trying to understand what was happening with the neural network, other members of the team were data wrangling new datasets. Below is the image of our first model.



We decided to try tensor flow to create our model due to the failure of the neural network. In our initial approach, we were given a similar result to the neural network. Because of this, we came to realize that our data that we were using was simply not enough for what we wanted to achieve. We happened upon a dataset that gave educated guesses on the global population at different years, dating all the way back to -10,000 B.C.. [3] This is the final dataset that we used for our project because it had a significantly more data for our model to improve upon and ultimately create the sigmoid shape of our model. Because of our choice in a sigmoid model, we wanted to have the graph of our population level out as it approached the maximum population that Earth can sustain, which predictions have said is approximately 13 - 15 billion. [4]

To start making predictions, we began by creating a loss equation to help guide our model. We used 7 variables: C1 through C7, which allows our model to change the shape and size of the sigmoid. When we first started training it, we were presented with a model that had stretched itself too far upward, with a result that would predict the population to be over 40 billion in 2122. This was obviously not ideal, and we realized that the problem with the model was that a single variable (C2) was taking over the model. This was because as the model was stretched higher and higher, the loss calculation became less and less. So, to fix this, we decided to add to the loss function a parameter that essentially punished the model for making any significant deviations from 1 on that variable. After implementing this, we were given a much more reasonable model that capped the curve in the next ~100 years at near 11 billion people. There was still some tweaking to be done, but we were confident that we were headed in the right direction.

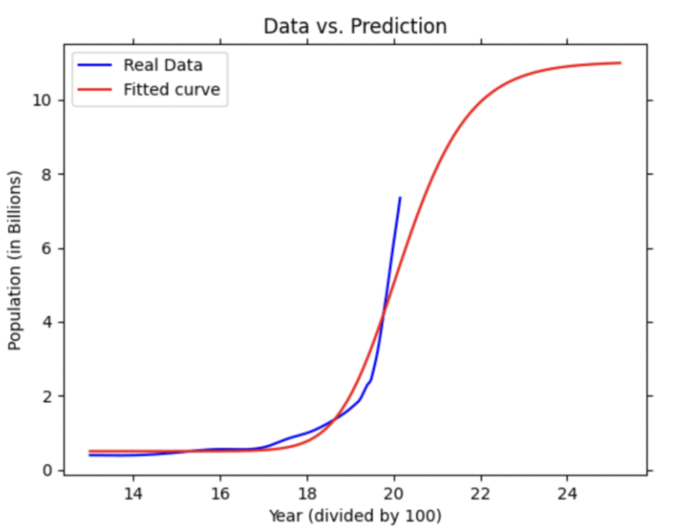
After a while of testing with these variables, we realized that our model that resulted was not exactly what we wanted. Looking at the graph, we were not satisfied with how closely our line was fitting the data even though the end prediction is in the range we were hoping for. The slope was not as steep as the actual data between the years of 1850 and 2000, and we wanted to make it fit better. So, in the final form, we had 8 variables, C1…C8 in our prediction equation which allowed the loss function to change the slope of the middle portion of the graph. The following equations are our prediction and loss equations used to tune our model.

$$\boxed{Prediction = \frac{c_1}{\left(c_2 + c_3\, e^{-c_4 (c_5 * Year + c_6)}\right)^{c_7}} + c_8}$$

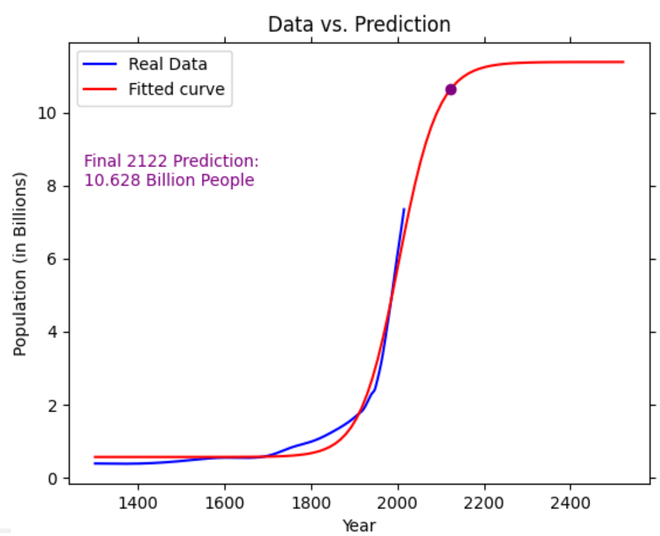
$$\boxed{Loss = \frac{\sum_i (labels - preds)^2}{batchSize} + 2 * (1 - self.c2)^2}$$

This resulted in a much better fit with a loss of less than 1%, and a prediction of ~10.6 billion people in 2122. View the following comparison to see the changes in the slope and fit of our curve on the real data provided to the model. There is a clear improvement towards the fitting of the curve while maintaining a similar prediction value.

BEFORE



AFTER



Overall this project forced us to rethink how we need to approach this problem multiple times. First, we struggled to make accurate predictions (we think) 100 years into the future using only data from the last century. Secondly, once we had found a dataset with estimates of previous historical population we were forced to consider how other factors outside of previous data could affect the curve, i.e. fertility and death rates and their trends. Lastly, on multiple occasions we had to edit the parameters of our model to reflect that of a sigmoid curve, all the while making sure it fit our data in a reasonable manner. Lots of credit to Gabriel for his expert knowledge on machine learning, neural networks, and the math behind why a good loss function is crucial to creating a well tuned model.

We are overall quite happy with the outcome of our model and the prediction value we received. Other projections from many online sources similarly suggest that Earth's population will grow beyond 10 Billion people in the next century but likely plateau as it becomes more common and easier to have less kids all the while as life expectancy around the globe rises. Pew Research indicates that Africa is one of the only continents that will continue to grow at a near 2.5% into the next 100 years, meanwhile China's population is expected to decline (and we can already see how some governments will play a role in this limitation to population growth). [5] Overall, we do not expect population growth to continue at such a significant rate and the factors leading to this helped us choose why a sigmoid curve is the best approximation model for the future population of Earth in the year 2122.

References

1. Fertility rate data: https://ourworldindata.org/world-population-growth
2. Death rate source: https://data.worldbank.org/indicator/SP.DYN.CDRT.IN
3. Dataset source: https://ourworldindata.org/world-population-growth
4. Earth sustainability: How many people can Earth actually support? (science.org.au)
5. Pew Research: The countries projected to gain and lose the most people by 2100 | Pew Research Center