Hack Climate Hackathon

sponsored by yOBI

Machine learn effect of Monsoon variability on India’s GDP

Mayank Rajoria
Prakhar Agrawal
Shubham Rathi

Roadmap

- Goals
- What did we model?
- Data Sources
- Methodology
- Comparing Outputs (No Monsoon data)
- Comparing Outputs (With Monsoon data)
- Concrete Math
- Conclusions

Goal

- **Aim:**
  - Find the effect of climate change on India's GDP
  - More simply put: Estimate future GDP based on monsoon and other standard factors such as Inflation, exchange rate etc.

What did we try to model?

MODEL EFFECT OF CLIMATE CHANGE ON THE FUTURE GDP

- Given parameters like population, inflation, unemployment, etc -> create a robust model to predict GDP of a country.
- Idea has its roots in **Gaussian Discriminant Analysis**
  - extend it to predict continuous values
  - unlike the original GDA which works only for classification problems
- Lateron try to incorporate the effect of monsoons
  - use the future weather forecasts and predict their effect on the GDP!

Data Sources

- **GDP:** https://github.com/Bhavikshah406/GDP-Prediction/blob/master/GDP_and_Major_Industrial_Sectors_of_Economy_Dataset.xls
- **Inflation:** http://www.inflation.eu/inflation-rates/india/inflation-india.aspx
- **Unemployment, real interest rate, CO2 emission, Consumption, Population:** http://data.worldbank.org/indicator/
- **Population:** http://www.worldometers.info/world-population/india-population/
- **Disasters:**
  - http://www.ndma.gov.in/en/disaster-data-statistics.html
  - http://aasc.nic.in/course%20material/disaster/documentation%20on%20past%20disasters.pdf
- **Exchange rate:** http://forecast-chart.com/usd-indian-rupee.html
- **Oil Price:** https://inflationdata.com/Inflation/Inflation_Rate/Historical_Oil_Prices_Table.asp
- **M1, M3 (Narrow-Broad) Money Value:** https://data.oecd.org/money/
- **Imports, Exports:** https://data.gov.in/catalog/exports-imports-and-trade-balance

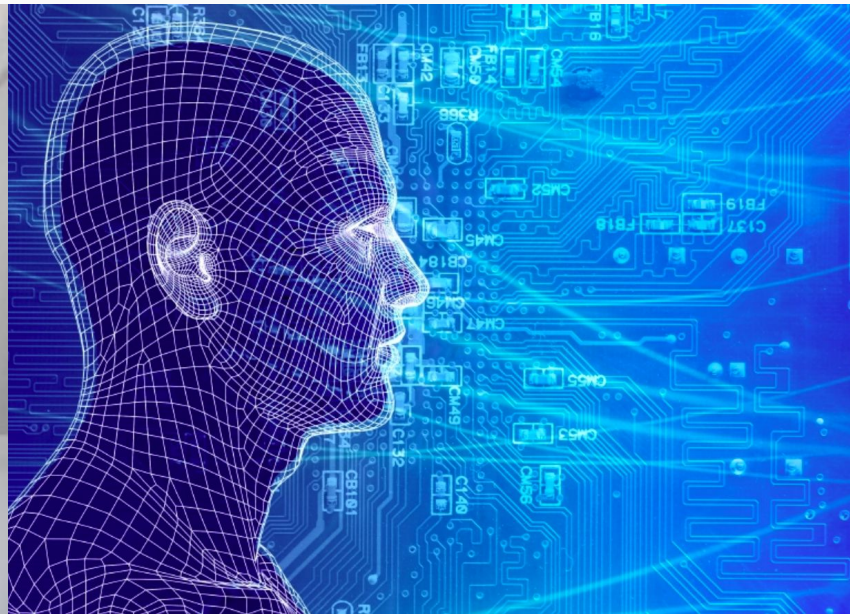
Methodology

- Data availability: Low (only ~60 years of data)
- Idea: Use a generative model to create a larger dataset
  - Something similar to **GDA** looked like a good approach
- Why use a generative model?
  - Discriminative models performed badly on the dataset
  - Could not properly model the data, given the small size of the dataset
  - Generative models found to performed better (as explained on next slide)
- Details:
  - Assumption: The complete dataset follows a **multivariate Gaussian**
  - Use **log likelihood** to arrive at a mean and variance
  - Use these parameters and the partially given dataset to predict the **target values**

Model Selection

- Other approaches tried:
  - **3-layer neural network** to predict the GDP.
  - **Ridge regression** with mean normalised dataset.
- Performance Evaluation:
  - Insufficient data to train the discriminative models
  - GDA type models eventually found to perform better





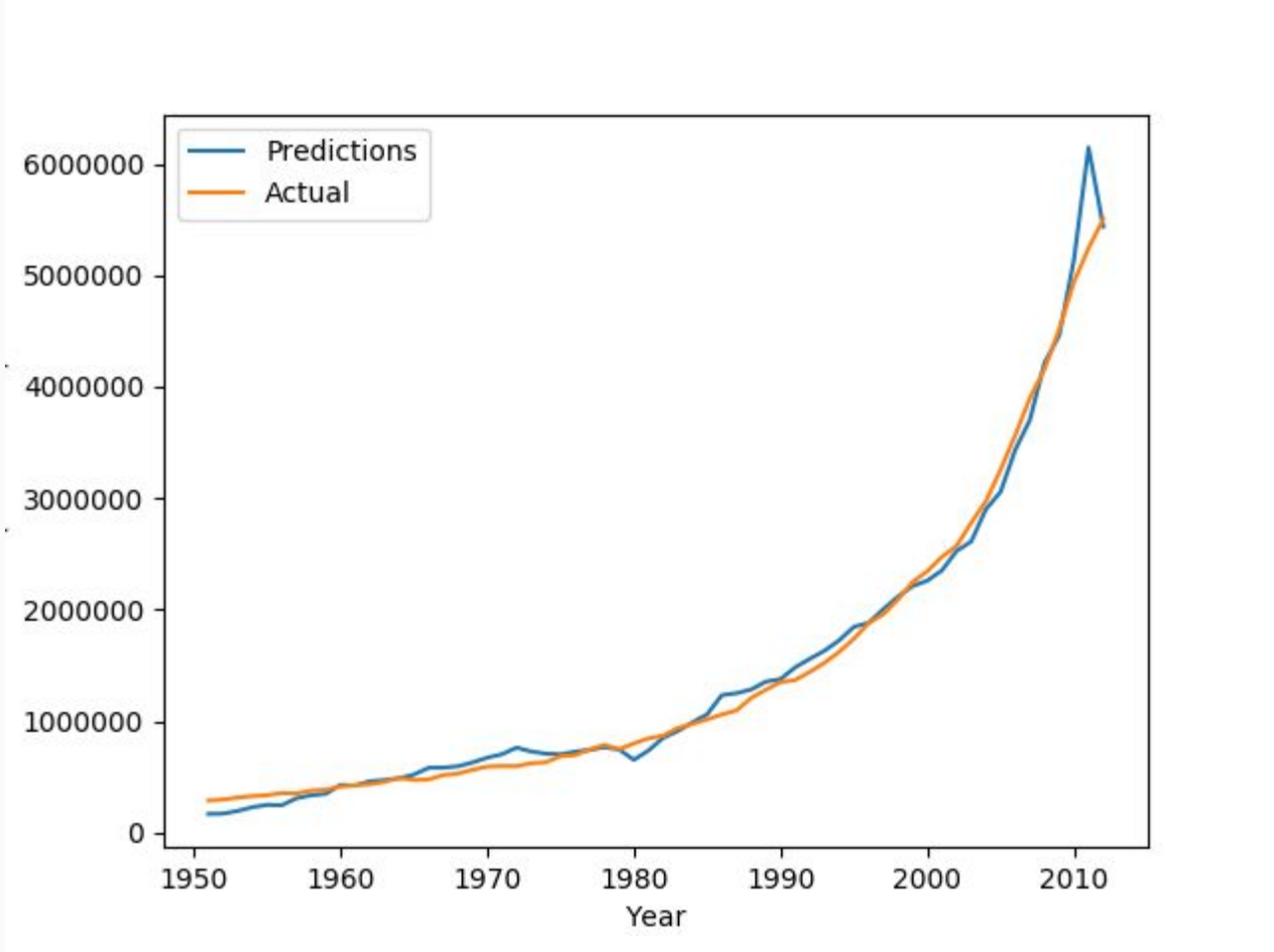
Comparing Outputs (No Monsoon data)

- **Output Comparison:**
  - Learning performed on 80% dataset
  - Prediction on rest 20%
  - Few data points from the validation set are shown below

Expected	Predicted	Ratio
286663.549904	286147.0	1.0018051907
294131.540714	294267.0	0.99953967218
311464.385554	312177.0	0.99771727434
324803.993045	325431.0	0.99807330292

Mean Absolute Error: 813.606730271

Comparing Outputs (No Monsoon data) 1



Prediction without monsoon and previous year's GDP

R2:

0.9884301925805549 8

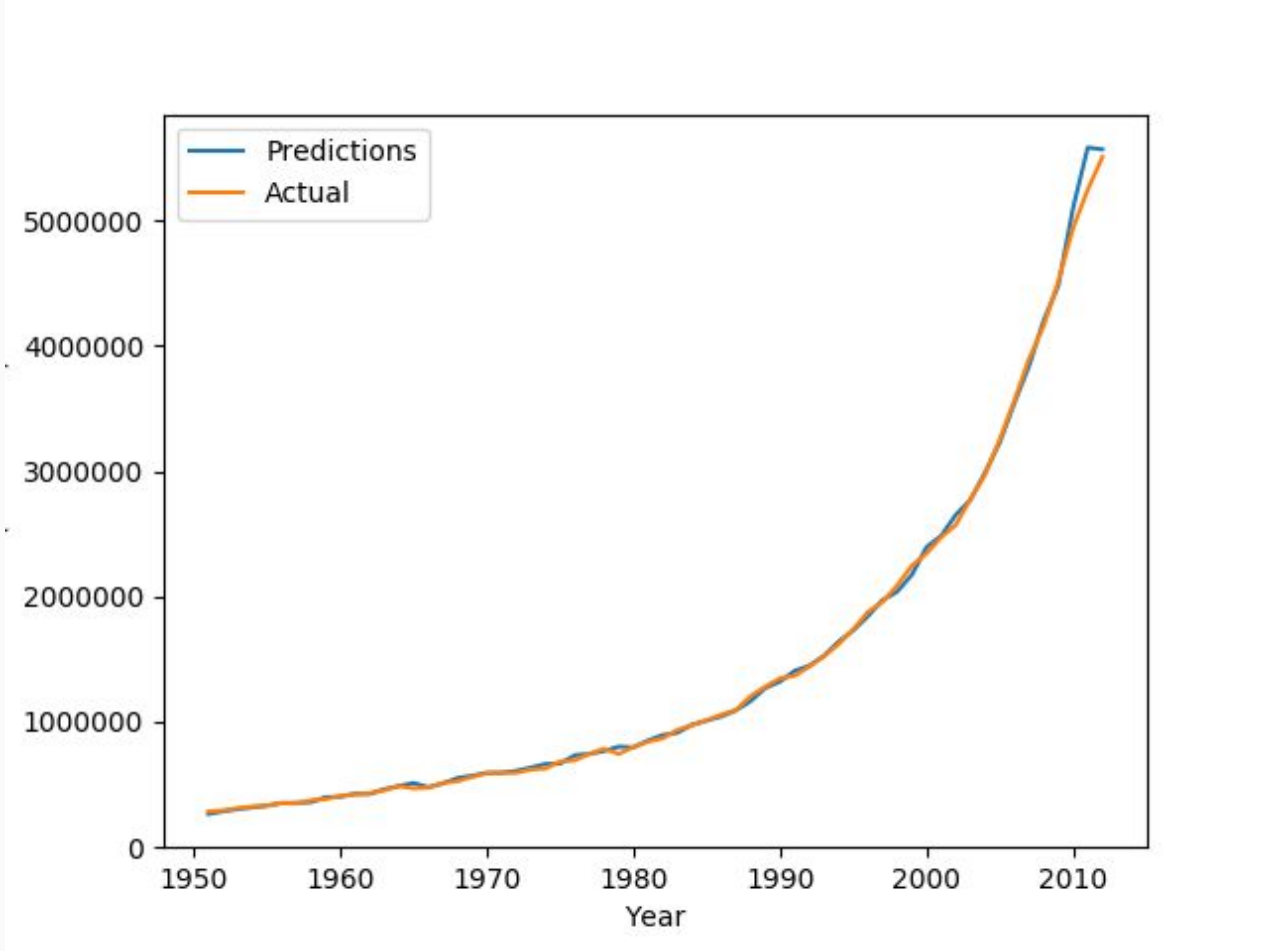
Std. Deviation of error:

146844.11254760568

Mean Absolute Error:

90825.028974907196

Comparing Outputs (No Monsoon data) 2



Prediction without monsoon but with previous year's GDP

R2:

0.9983936677573766

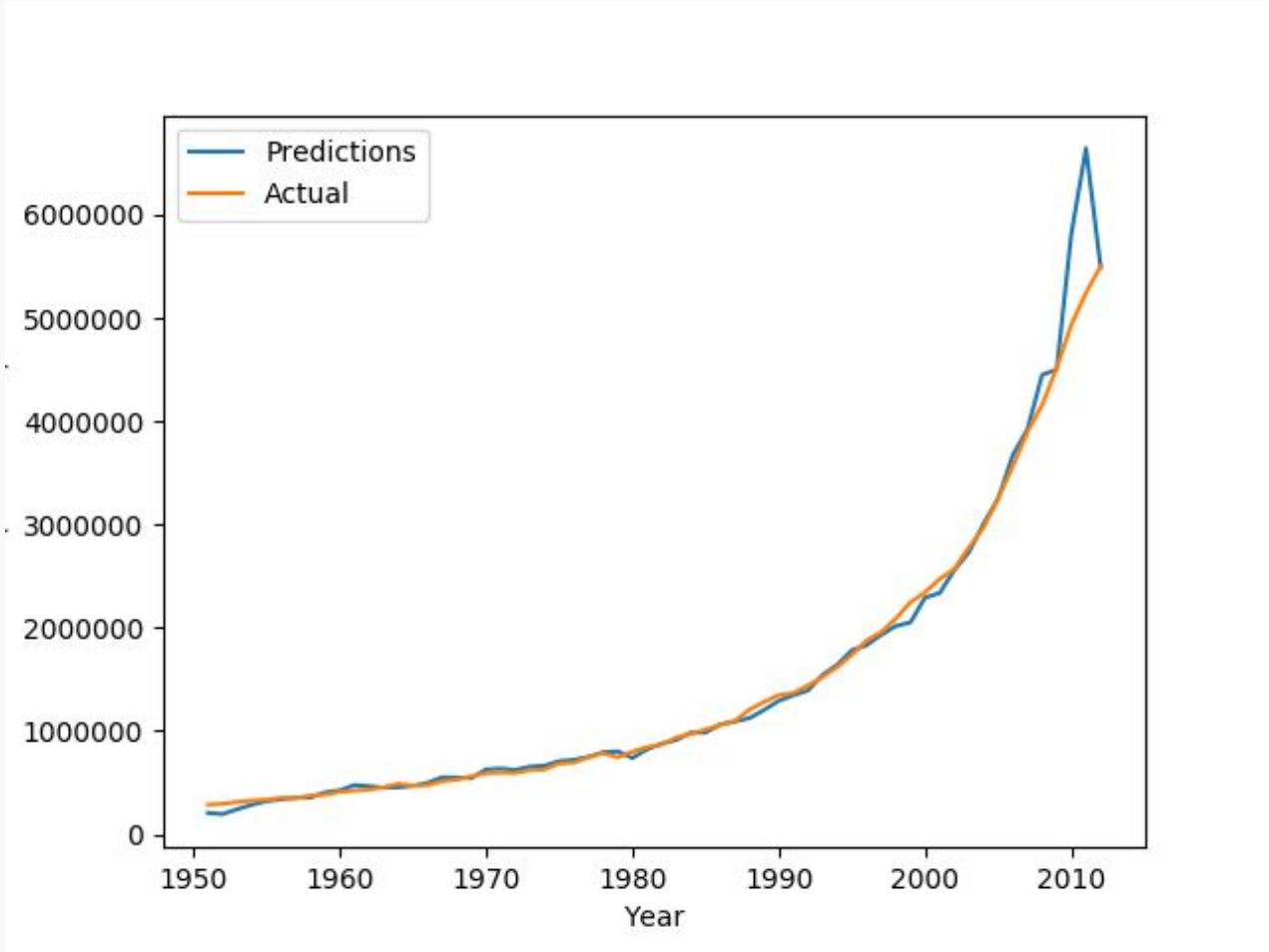
Std. Deviation of error:

53665.916501780252

Mean Absolute Error:

27571.770762905719

Comparing Outputs (With Monsoon data) 3



Prediction with monsoon but without previous year's GDP

R2:

0.9775356560364546 5

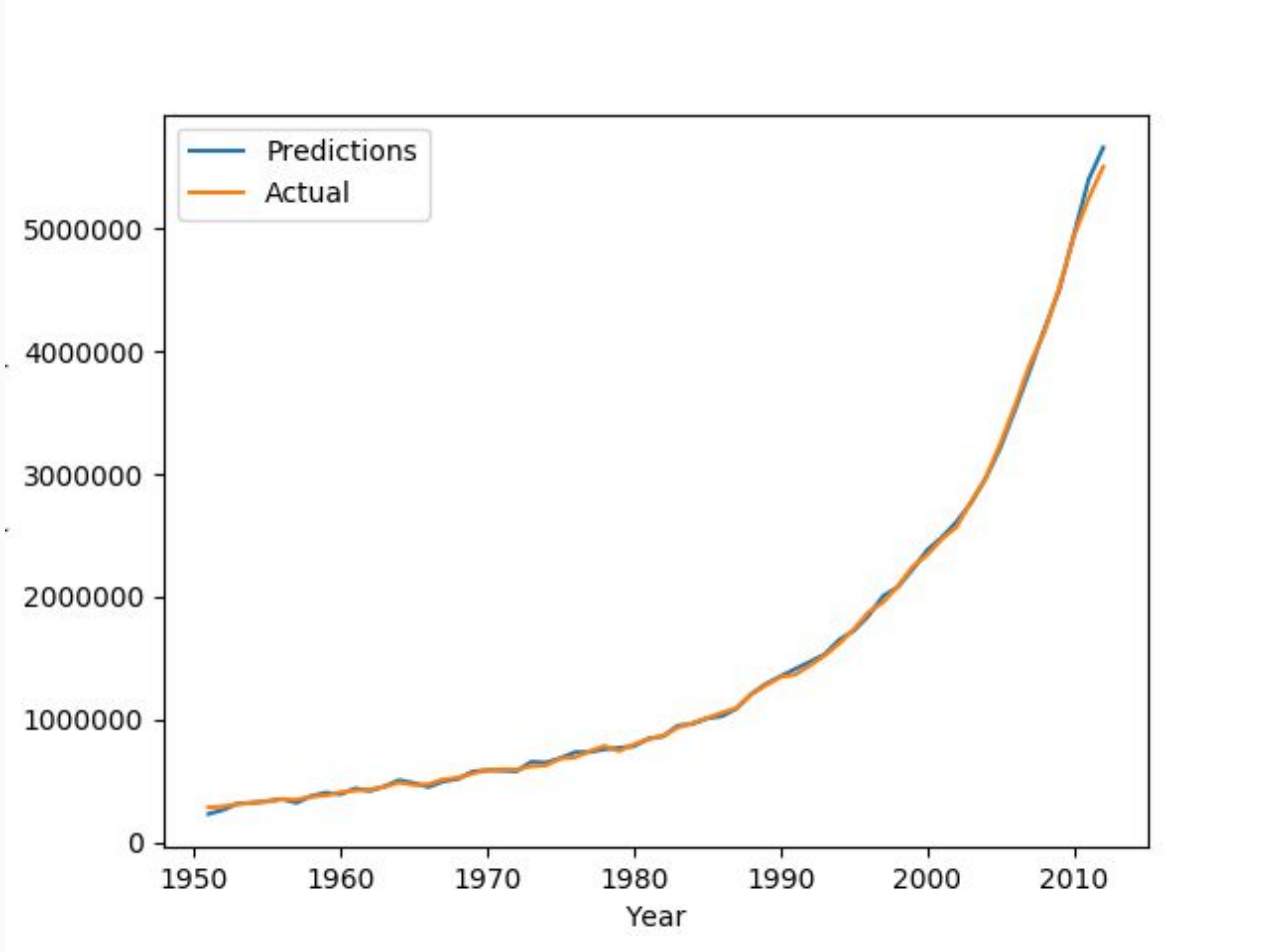
Std. Deviation of error:

216726.54775110719

Mean Absolute Error:

77038.194316786306

Comparing Outputs (With Monsoon data) 4



Prediction with monsoon and previous year's GDP

R2:

0.9992991627346228 3

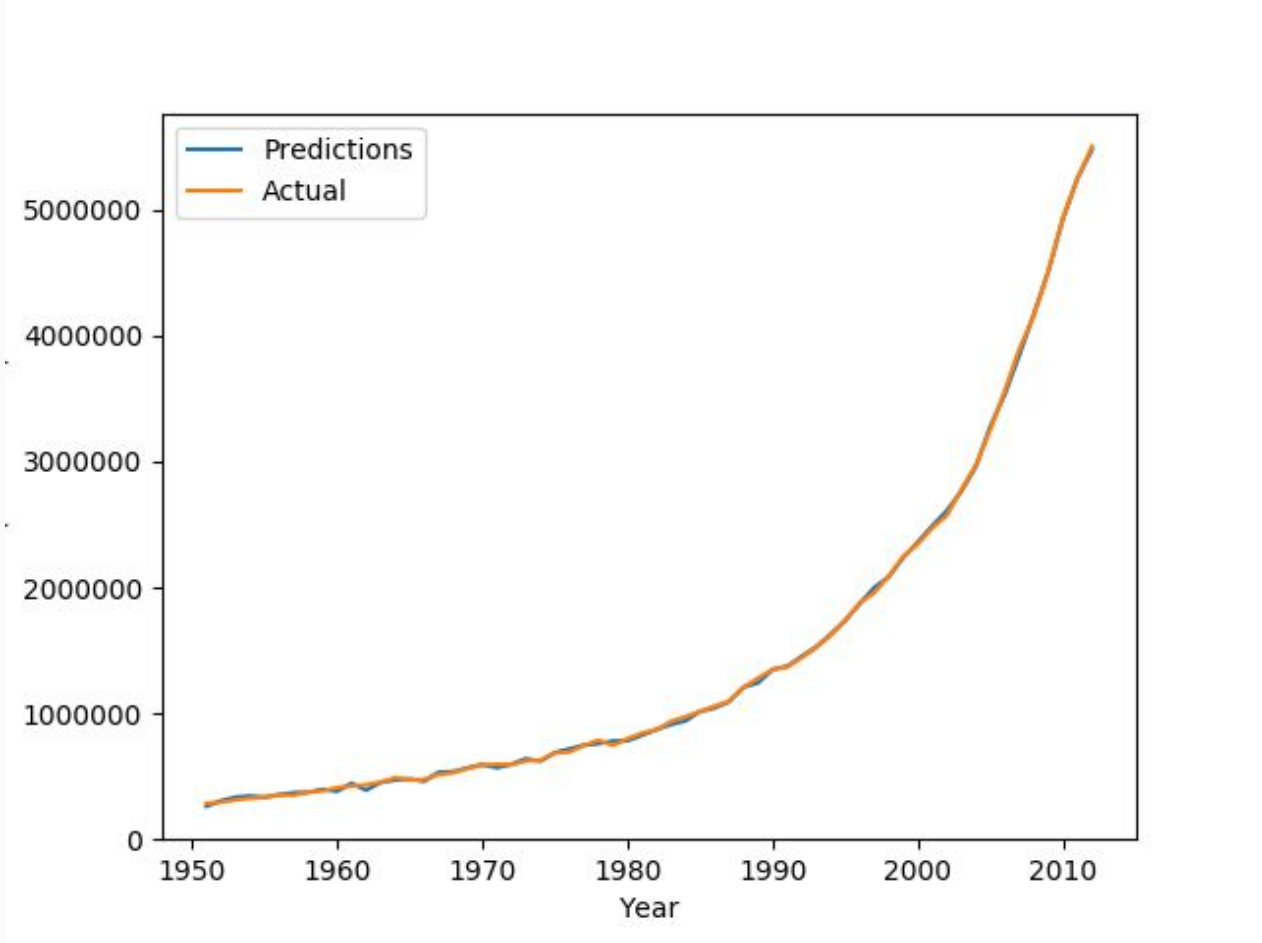
Std. Deviation of error:

35361.194412215729

Mean Absolute Error:

22560.337999582905

Comparing Outputs (With Monsoon data) 5



Prediction without Imports and Exports

R2:

0.9997895105005524 3

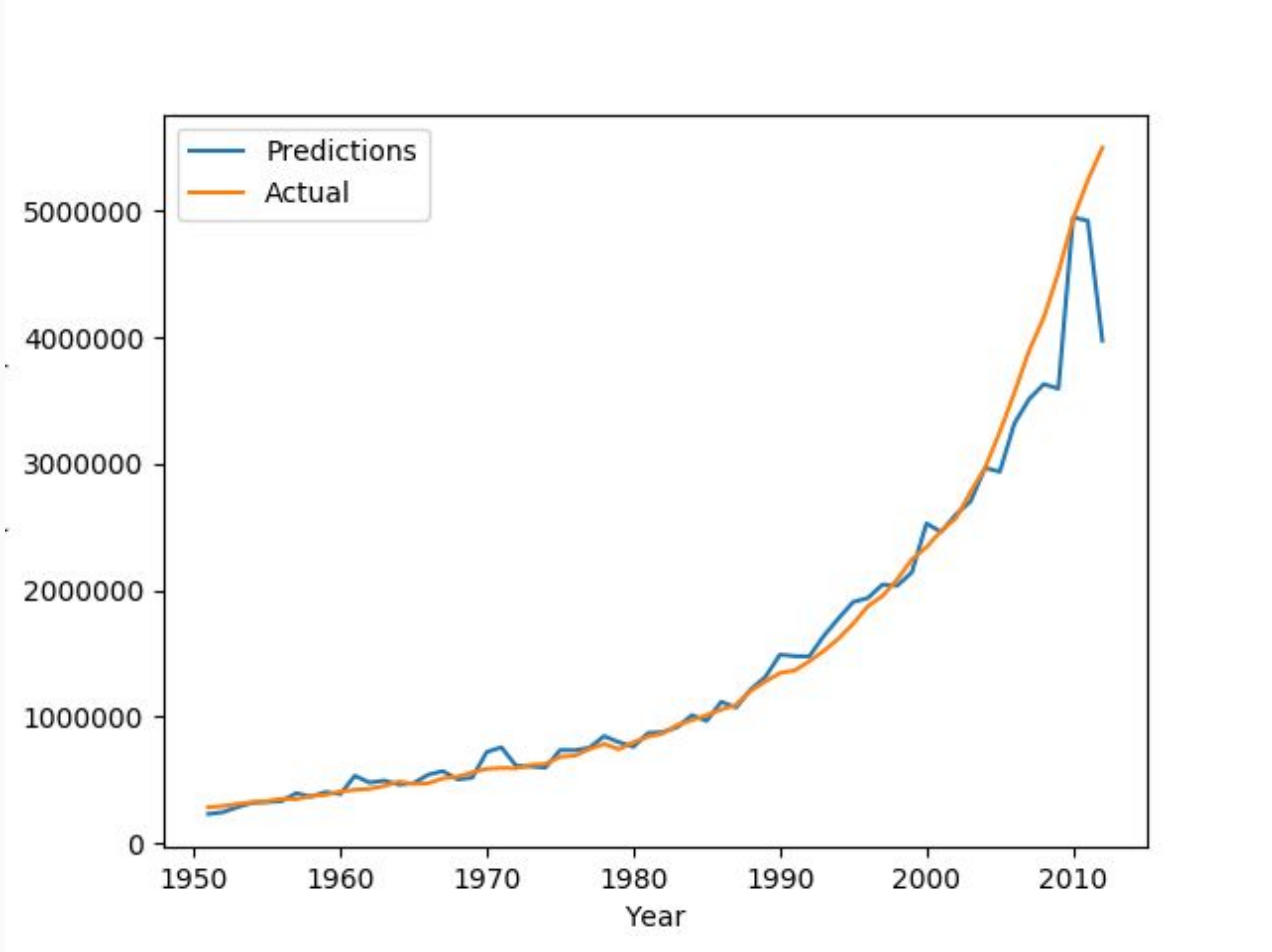
Std. Deviation of error:

19370.622864863744

Mean Absolute Error:

15166.41110596468

Comparing Outputs (With Monsoon data) 6



Prediction without Narrow Money and Broad Money

R2:

0.9525566184507803

Std. Deviation of error:

254867.9073921626

Mean Absolute Error:

115546.96024132275

Correlation of GDP with parameters

Correlation of GDP with:-

- Inflation	:	0.1224586492
- Population	:	0.9051026464
- Consumption	:	-0.8858570584
- Disasters	:	0.602740577
- Nominal Oil Prices	:	0.9029063566
- Inflation adjusted Oil Prices	:	0.5309821159
- Narrow Money	:	0.9680957827
- Broad Money	:	0.9488423126
- CO2 Emissions	:	0.9727361439
- Exports	:	0.9403667282
- Imports	:	0.9233833463
- Previous Year’s GDP	:	0.999619752

Correlation of GDP with rainfall parameters

Correlation of GDP with:-

- Rainfall in		
- Q1	:	-0.05883275104
- Q2	:	0.06134254094
- Q3	:	-0.1541715853
- Q4	:	-0.1201921931
- Rainfall in previous Q3	:	-0.1704237816
- Maximum monthly rainfall	:	-0.240898835
- Minimum Monthly Rainfall	:	-0.007562645287
- Total yearly rainfall	:	-0.1299957776

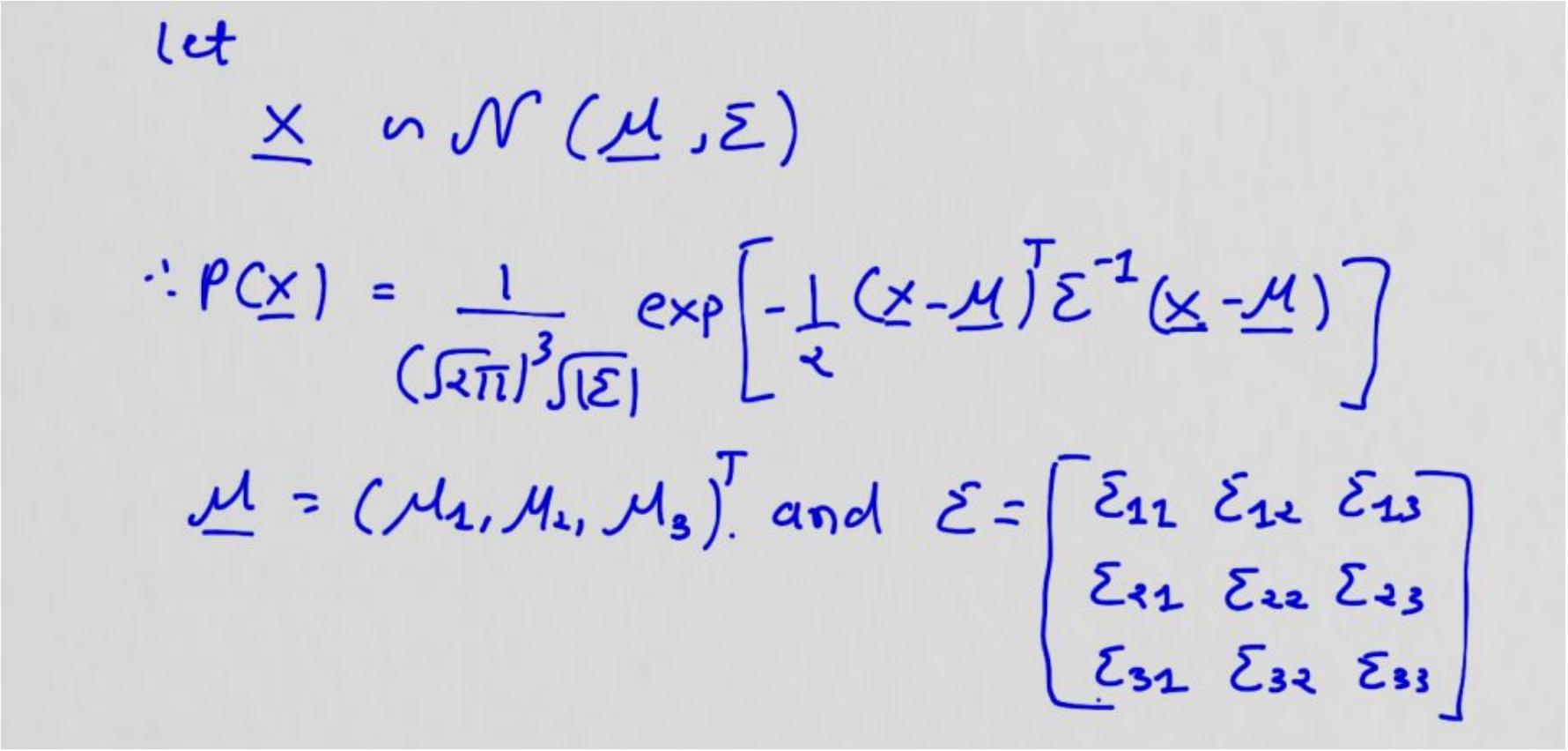
Correlation between yearly difference of GDP and yearly difference of rainfall:

0.10511335

*where Q indicates Quater

Concrete Math: Model

- We show the math for a simple case of predicting X1, given X2 and X3 where X1,X2,X3 are random variables.
- Let X = (X1,X2,X3)
- Assumption: The joint distribution follows closely with a **multivariate Gaussian**



Concrete Math: Learning

- Data Likelihood:

Log-likelihood of training data

$$LL(\underline{\mu}, \Sigma) = \sum_{i=1}^{m} \log P(\underline{x}^{(i)})$$

$$LL(\underline{\mu}, \Sigma) = \sum_{i=1}^{m} \left[-\log (2\pi)^{3/2} - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\underline{x}^{(i)} - \underline{\mu})^T \Sigma^{-1} (\underline{x}^{(i)} - \underline{\mu}) \right] \quad \text{— (1)}$$

- Learning Parameters:

Mean ->

$$\nabla_{\underline{\mu}} LL(\underline{\mu}, \Sigma) = 0$$

$$\Rightarrow \boxed{\underline{\mu} = \frac{1}{m} \sum_{i=1}^{m} \underline{x}^{(i)}}$$

$$\nabla_{\Sigma} LL(\underline{\mu}, \Sigma) = 0$$

$$\Rightarrow \boxed{\Sigma = \frac{1}{m} \sum_{i=1}^{m} (\underline{x}^{(i)} - \underline{\mu})(\underline{x}^{(i)} - \underline{\mu})^T}$$

<- Covariance Matrix

Concrete Math: Prediction

Then our Gaussian distribution $P(x)$ can be rewritten as,

$P(Y_1, Y_2)$

$\mu' = [\mu_1' \ \mu_2']^T$ and $\Sigma' = \begin{bmatrix} \Sigma'_{11} & \Sigma'_{12} \\ \Sigma'_{21} & \Sigma'_{22} \end{bmatrix}$

Conditioning on Y_2, we have distribution $P(Y_1 / Y_2 = y_2)$ with

$$\mu'_{1|2} = \mu_1' + \Sigma'_{12} \Sigma'^{-1}_{22} (y_2 - \mu_2')$$

Now,

$$E(Y_1 | Y_2 = y_2) = \mu'_{1|2}$$

Conclusions

- Monsoon and GDP may show a **weak correlation** within a year, and a very low or negligible correlation over a longer span.
- Correlation Coefficient: **-0.1541** (over ~60 years of data)
- Our model is another confirmation to the claims in the past papers, stating low or negligible relation between monsoon and GDP of a country
- These papers are related below:
  - https://www.jstor.org/tc/accept?origin=/stable/pdf/4418949.pdf
- We could successfully predict the GDP with high accuracy as indicated in the results section of the presentation.

Conclusions

- The primary indicators for the GDP calculation from our analysis come out to be:
  - **Population**: With an increase in population, there will be an increased value of goods and services produced and consumed, causing the observed increase in GDP.
  - **CO_2 Emissions:** CO_2 might be indirectly reflection the production of goods, energy, factories etc in the country, hence having a high positive correlation with GDP.
  - **M1** (Narrow Money), **M3** (Broad Money)
  - **Imports/Exports**: These are the transaction happening between India and other country and are used in the calculation of GDP, hence there is high correlation.
  - **Oil Prices**: Oil is a major commodity that affects a country's economy and hence affects GDP as expected

Thank you :)

The source code, dataset and other files are available on the github repo
https://github.com/Prakhar0409/GDP-Prediction-Engine