```
My Code for the task can be viewed at:
https://github.com/sayarghoshroy/COVID-19_Modelling_for_India
```

1) Problem: We model the coronavirus outbreak statistics for countries & train our model for India. Based on all the provided references, I decided to use the SIR model.

Data: The dataset had the following fields: new_daily_cases, new_daily_deaths, total_deaths, total_recoveries, total_cases for a span of 94 dates. Using the raw data, we could calculate: infected_present = total_cases − total_deaths − total_recoveries & susceptible_present = N − total_cases for each particular day. Now; susceptible_present, infected_present, and total_recoveries serve as the gold standard S, I, R values. All relevant plots were generated to get a proper feel of the dataset. SIR model is a good choice given the nature & trends in data. Though N is supposed to be population size, putting N as 1.5 billion was not feasible as it made everything else statistically insignificant. Therefore, I set $N \sim 10^5$ and used the intuition of 'exposed_directly' as a feature to select 'N' subset of population for our modelling.

Now, by SIR model:

$$\frac{dS}{dt} = -\frac{\beta S I}{N}, \quad \frac{dI}{dt} = \frac{\beta S I}{N} - \gamma I, \quad \frac{dR}{dt} = \gamma I$$: we train to find β & γ.

Now, let $\bar{S}, \bar{I}, \bar{R}$ be the computed S, I, R values using the model for time = 1 to d (d being the number of days i.e 94). We define loss $\mathcal{L}$ = L2 norm of (infected_present − $\bar{I}$). Denoting infected present as I.

$$\mathcal{L} = [\bar{I} - I]^T [\bar{I} - I]$$

$$\therefore \frac{\partial \mathcal{L}}{\partial \beta} = 2 \cdot [\bar{I} - I] \cdot \frac{\partial \bar{I}}{\partial \beta} = 2[\bar{I} - I] \cdot \frac{\bar{I}(\beta + \Delta\beta) - \bar{I}(\beta)}{\Delta\beta}$$

$$\& \ \frac{\partial \mathcal{L}}{\partial \gamma} = 2[\bar{I} - I] \cdot \frac{\partial \bar{I}}{\partial \gamma} = 2[\bar{I} - I] \cdot \frac{\bar{I}(\gamma + \Delta\gamma) - \bar{I}(\gamma)}{\Delta\gamma}$$

- $\Delta\beta$, $\Delta\gamma$ being very small values.
- We integrate over the range of $t = 0$ to 94 on the given differentials to generate all $\bar{I}$ values.
- Note; $\frac{\partial \bar{I}}{\partial \beta}$ cannot be computed analytically. This procedure gives us a fairly accurate approximation.

Now, we have: $W = [\beta, \gamma]^T$. Update for simple GD.

$$W^{(k+1)} \leftarrow W^{(k)} - \eta \cdot 2 \begin{bmatrix} (\bar{I} - I) \cdot \frac{\bar{I}(\beta + \Delta\beta) - \bar{I}(\beta)}{\Delta\beta} \\ (\bar{I} - I) \cdot \frac{\bar{I}(\gamma + \Delta\gamma) - \bar{I}(\gamma)}{\Delta\gamma} \end{bmatrix}$$

And for Newton's updates: Hessian computation:

$$H = \begin{bmatrix} \partial^2 L/\partial\beta^2 & \dfrac{\partial^2 L}{\partial\beta\,\partial\gamma} \\ \dfrac{\partial^2 L}{\partial\gamma\,\partial\beta} & \partial^2 L/\partial\gamma^2 \end{bmatrix}.$$

And $$\frac{\partial^2 L}{\partial\beta^2} = 2\left[(\bar{I}-I)\frac{\partial^2 \bar{I}}{\partial\beta^2} + \left(\frac{\partial \bar{I}}{\partial\beta}\right)^2\right]$$

where $$\frac{\partial^2 \bar{I}}{\partial\beta^2} = \frac{\bar{I}(\beta+2\Delta\beta) + \bar{I}(\beta) - 2\,\bar{I}(\beta+\Delta\beta)}{(\Delta\beta)^2}$$

$$\frac{\partial^2 L}{\partial\beta\,\partial\gamma} = \frac{\partial^2 L}{\partial\gamma\,\partial\beta} = 2\left[(\bar{I}-I)\frac{\partial^2 \bar{I}}{\partial\gamma\,\partial\beta} + \left(\frac{\partial\bar{I}}{\partial\beta}\right)\left(\frac{\partial\bar{I}}{\partial\gamma}\right)\right]$$

where $$\frac{\partial^2 \bar{I}}{\partial\beta\,\partial\gamma} = \frac{\bar{I}(\beta+\Delta\beta,\gamma+\Delta\gamma) - \bar{I}(\beta,\gamma+\Delta\gamma) - \bar{I}(\beta+\Delta\beta,\gamma) + \bar{I}(\beta,\gamma)}{(\Delta\beta)(\Delta\gamma)}$$

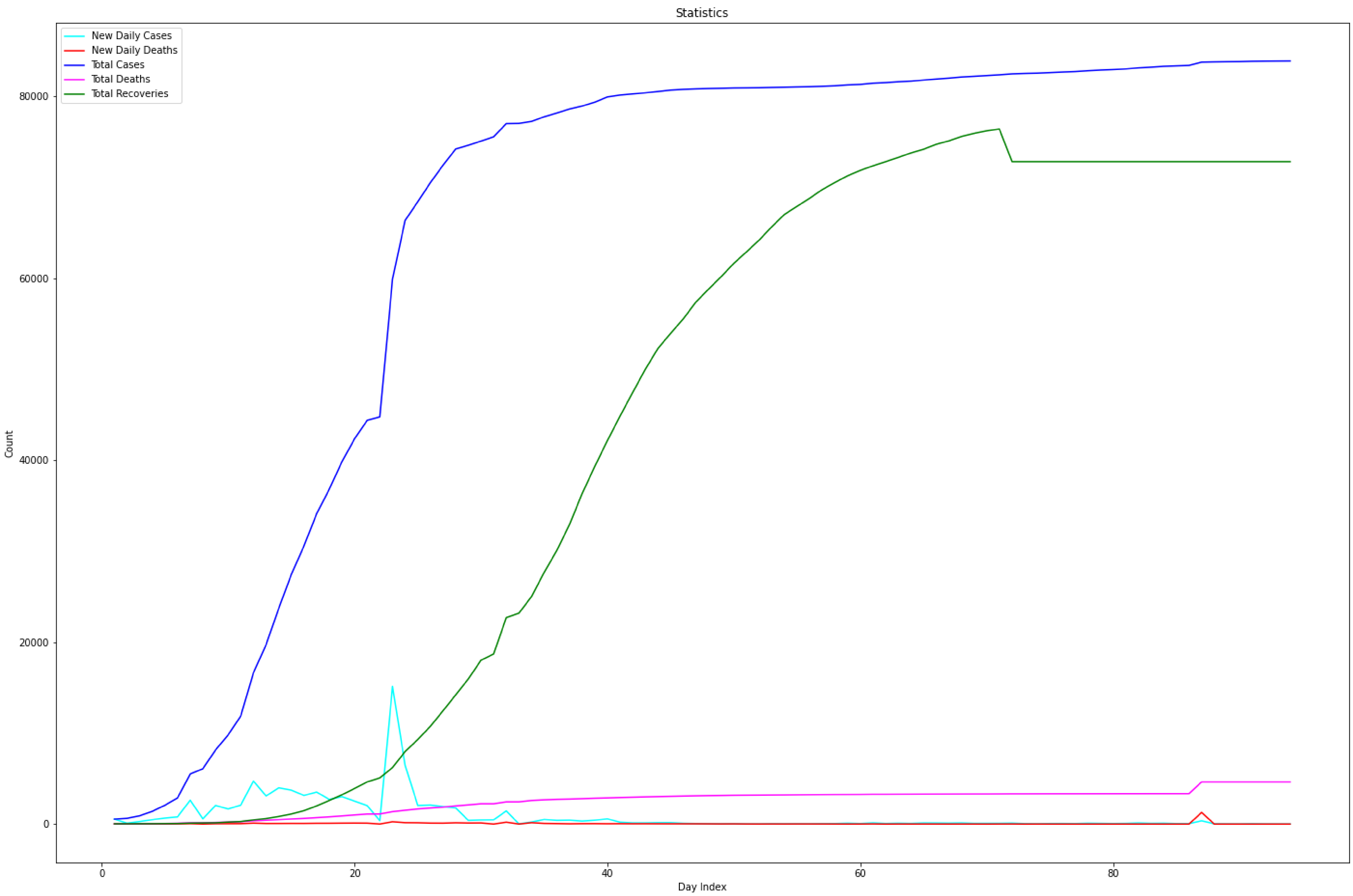
$$\frac{\partial^2 L}{\partial\gamma^2} = 2\left[(\bar{I}-I)\frac{\partial^2\bar{I}}{\partial\gamma^2} + \left(\frac{\partial\bar{I}}{\partial\gamma}\right)^2\right]$$

where $$\frac{\partial^2\bar{I}}{\partial\gamma^2} = \frac{\bar{I}(\gamma+2\Delta\gamma) + \bar{I}(\gamma) - 2\,\bar{I}(\gamma+\Delta\gamma)}{(\Delta\gamma)^2}$$

$\therefore$ Update with Newton's Method:

$$w^{(k+1)} \leftarrow w^{(k)} - H^{-1}\begin{bmatrix} (\bar{I}-I)\dfrac{\bar{I}(\beta+\Delta\beta) - \bar{I}(\beta)}{\Delta\beta} \\ (\bar{I}-I)\dfrac{\bar{I}(\gamma+\Delta\gamma) - \bar{I}(\gamma)}{\Delta\gamma} \end{bmatrix}$$

- We run GD with random initialization + random restarts to compute $w^* = [\beta^*, \gamma^*]^T$.
- The η is learn rate. Results show that use of a scaling factor with Newton's updates speeds up the learning process.


Statistics
New Daily Cases
New Daily Deaths
Total Cases
Total Deaths
Total Recoveries
Count
0
20000
40000
60000
80000
Day Index
0
20
40
60
80