E0 259 Data Analytics
COVID-19 Assignment

Dataset description:
The CSV file (COVID19_data.csv) contains three time series for the state of Karnataka.

1. The cumulative number of reported cases (column name: confirmed) until 20 September 2021.
2. The cumulative number of tests done (column name: tested) until 20 September 2021.
3. The cumulative number of vaccinations (dose 1) administered (column name: First Dose Administered) since 16 January 2021 until 20 September 2021.

(The CSV file also contains a few extra columns like Recovered, Deceased, Other, Second Dose Administered and Total Doses Administered. These columns are not needed to complete this assignment. These are being provided for the students who want to experiment with advanced models.)

Problem description:

1. Implement the SEIRV model with immunity waning in a population of N individuals.
   (a)
$$\begin{aligned}
\Delta S(t) &= -\beta(t)S(t)\frac{I(t)}{N} - \varepsilon\Delta V(t) + \Delta W(t),\\
\Delta E(t) &= \beta(t)S(t)\frac{I(t)}{N} - \alpha E(t)\\
\Delta I(t) &= \alpha E(t) - \gamma I(t),\\
\Delta R(t) &= \gamma I(t) + \varepsilon\Delta V(t) - \Delta W(t).
\end{aligned}$$
   (b) Use the mean incubation period $\alpha^{-1} = 5.8$ days, mean recovery period $\gamma^{-1} = 5$ days, vaccine efficacy $\varepsilon = 66\%$, and the total population $N = 70$ million.
2. Calibrate the model parameters to match the daily COVID-19 reported cases in Karantaka during 16 March 2021 – 26 April 2021.
   (a) Use the following constraints on the initial conditions:
      i. $R(0)$ is between 15.6% and 36% of the population.
      ii. The initial cases-to-infections ratio (CIR(0)) is between 12.0 and 30.0.
   (b) Use the following model for immunity waning:
      i. Every recovered individual at time 0 becomes susceptible after 30 days. That is, set $\Delta W(t) = R(0)/30$, when t is between 16 March 2021 and 15 April 2021.
      ii. Immunity waning occurs after 180 days. That is, set $\Delta W(t) = \Delta R(t-180) + \varepsilon\Delta V(t-180)$, when t is larger than 11 September 2021.
   (c) Update the cases-to-infections ratio at time t as
$$\mathrm{CIR}(t) = \mathrm{CIR}(0)\frac{T(t_0)}{T(t)},$$
   where $T(t)$ is the average number of tests done during the past 7 days (i.e., during $t-7$ to $t-1$), and t_0 is 16 March 2021.

(d) Let $\Delta\text{confirmed}(t)$ denote the daily cases time-series. Take the running seven day average of $\Delta\text{confirmed}(t)$, from the spreadsheet, and call it $\bar{c}(t)$. Similarly, take the running seven day average of $\Delta i(t)$ and call it $\overline{\Delta i}(t)$.

(e) Use the loss function

$$l(P) = \frac{1}{42} \sum_{t=\text{March 16}}^{\text{April 26}} \left(\log(\bar{c}(t)) - \log(\overline{\Delta i}(t))\right)^2,$$

where $P = (\beta, S(0), E(0), I(0), R(0), \text{CIR}(0))$ are the unknowns and $\Delta i(t) = \frac{\alpha E(t)}{\text{CIR}(t)}$.

(f) Use the gradient descent algorithm to minimise this loss function over the unknowns. That is, implement the updates

$$p(j+1) = p(j) - \frac{1}{j+1}\partial_p l(p(j)),\ p \in P,\ j \geq 0,$$

subject to the conditions (a)–(c) above. Stop at step n when $l(P(n)) \leq 0.01$.

To estimate the gradient perturb β on either side by ± 0.01, perturb $\text{CIR}(0)$ on either side by ± 0.1, and all other parameters by ± 1, and estimate the gradient.

3. Use 2 lakhs vaccinations per day starting 27 April 2021 and run the predictions forward until 31 December 2021 for the following scenarios.

   (a) Open loop controls: Use the contact rate β obtained in Step 2 throughout the duration of prediction. Repeat this with the contact rates $2\beta/3$, $\beta/2$, and $\beta/3$.

   (b) Closed loop control: Starting each Tuesday, compute the average number of new cases in the preceding one week, and use the contact rate for current week as shown in Table 1.

Average number of new cases in the previous week	Contact rate for the current week
< 10000	β
$10001 - 25000$	$2\beta/3$
$25001 - 100000$	$\beta/2$
> 100001	$\beta/3$

Table 1: Closed loop control

4. In your report, generate a plot that shows the number of new cases predicted on each day in all the five scenarios above. In the same plot, include the number of new reported cases in Karnataka until 20 September 2021. Also, generate a plot that shows the evolution of the fraction of the susceptible population in all the five scenarios.

Submit your code and report as a single .zip file that contains as a single .py file and a single .pdf file. The code should run with the csv file picked from the parent directory, namely, ../COVID19_data.csv.