AML Programming Assignment I: Gaussian Process for Predicting Vaccinations

Name : Sunandan Adhikary

Roll : 21CS91R14

**** Please change the first column header *Day Number as Day_Number* in the .xlsx file**

Prerequisites and Execution :

- Execute 'pip install matplotlib scipy numpy pandas openpyxl' before execution
- Run 'python 21CS91R14_sunandan.py' in the directory of the provided python file

Methodology implemented:

1. Imported the x and noisy y values i.e. respectively day index and vaccination counts, as training data set from csv (noise variance given = 0.1).
2. Considered a 300 day index as test data x for which we need to predict y values i.e.daily and total vaccinations on x-th day.
3. Used a RBF Kernel Function considering the exponential nature of the growth of y values with constant zero mean for training data
    a) Created a Kernel Function kernel_func() to calculate different types of Kernels for input sets of data and for input kernel parameters.
    b) Created a parameter optimization function mleopt() to minimize the negative log likelihoods of the training y values given the x values and return the MLE parameters for the most fit kernel creation.
4. Calculated Kernels for each of the training and test data sets and their combinations (i.e. covariance of x values from different or same datasets)
5. Then by calculating the conditional mean and co-variance values (i.e. test y value for given test x values and training data) we get the posterior predicted distribution of the daily and total vaccination stats i.e. test y values.
6. We can see our resulted plots of expected daily vaccination (green) and expected total vaccinations (blue) are fitting the training data points (+ marked) well and variance (red area) is quite less due to large number of training data points (refer to

Figure1).

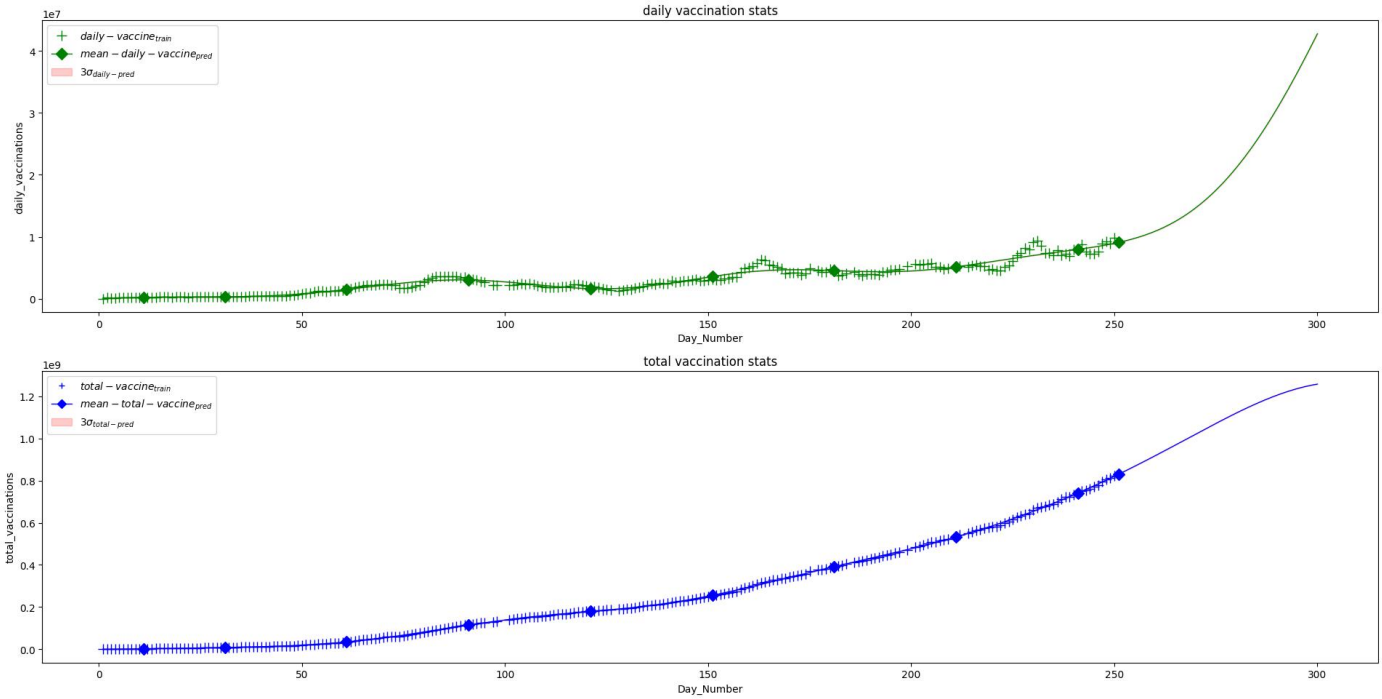

Figure 1 : PPD of Vaccination Statistics Using GP

7. Since the Vaccination counts are in the scale of 1e+8 , the uncertainty is almost not visible in the posterior mean-variance plot (Figure 1). Please refer to Figure2 to get an idea about how the uncertainty increases due to scarce training/observation data points. In Figure 2 we have plotted the daily and total vaccination stats in y axis in 1e-8 scale and x axis contains the day number. The red shaded regions around both daily and total prediction curves denote +-3-sigma uncertainty.

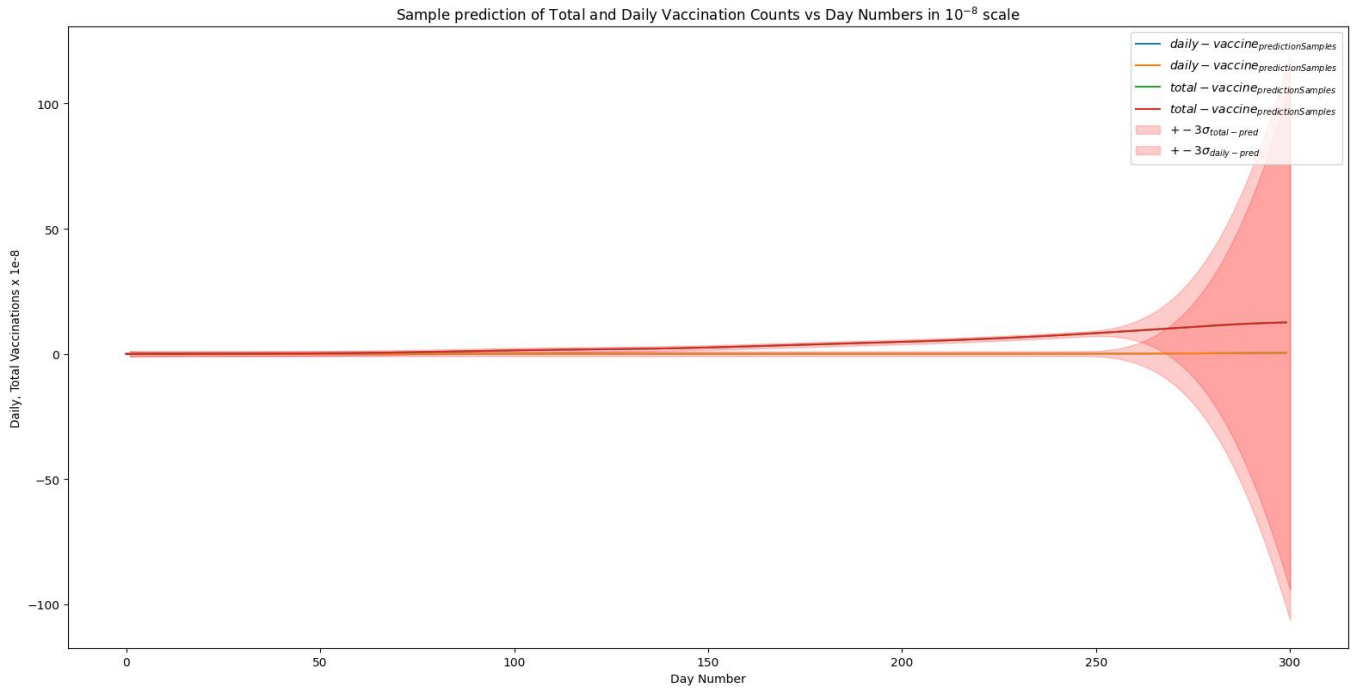

Figure 2 : Sample predictions with 6-sigma region in 1e-8 scale

8. Expected values of some missing/desired vaccination stats are diamond marked and presented below.

Results :

With optimized Parameter Values (for 300 predictions, with 240 training data), ***l = 51.34676895367992 and sigma_F = 184.44057888035516 for rbf kernel,*** expected (prediction of)

1 daily vaccination on 10-th day is :

242454.26626843945 with +-0.6561152965970674 deviation

and total vaccination on 10-th day is :

2115228.951679281 with +-0.6561152965970674 deviation

2 daily vaccination on 30-th day is :

321673.2351947574 with +-0.6488801839784513 deviation

and total vaccination on 30-th day is :

8124444.634388105 with +-0.6488801839784513 deviation

3 daily vaccination on 60-th day is :

1514854.7159018596 with +-0.6453822575329511 deviation

and total vaccination on 60-th day is :

34654849.15158034 with +-0.6453822575329511 deviation

4 daily vaccination on 90-th day is :

3044719.683409435 with +-0.6458731796667931 deviation

and total vaccination on 90-th day is :

114394686.90117808 with +-0.6458731796667931 deviation

5 daily vaccination on 120-th day is :

1652265.9870359118 with +-0.6449414775614183 deviation

and total vaccination on 120-th day is :

178992154.10151002 with +-0.6449414775614183 deviation

6 daily vaccination on 150-th day is :

3603789.543182654 with +-0.6446816271681943 deviation

and total vaccination on 150-th day is :

255544257.73722947 with +-0.6446816271681943 deviation

7 daily vaccination on 180-th day is :

4562594.698429172 with +-0.6461821679477411 deviation

and total vaccination on 180-th day is :

391507017.6112783 with +-0.6461821679477411 deviation

8 daily vaccination on 210-th day is :

5171810.447532794 with +-0.648205392195121 deviation

and total vaccination on 210-th day is :

532379260.83186066 with +-0.648205392195121 deviation

9 daily vaccination on 240-th day is :

7976422.651809405 with +-0.6565284682568353 deviation

and total vaccination on 240-th day is :

740535337.9393176 with +-0.6565284682568353 deviation

10 daily vaccination on 250-th day is :

9092291.942169132 with +-0.7861186473655716 deviation

and total vaccination on 250-th day is :

829543153.947675 with +-0.7861186473655716 deviation