

Programming Assignment 1: Gaussian Process

The objective of this assignment is to predict the number of vaccinations for COVID-19 in India using Gaussian Process Regression.

The CSV data file `vaccination.csv` contains three columns: day number, total_vaccination, daily vaccination, in India. The goal is to predict the number of daily vaccinations (column 3) in India and the world for a day number (in future). Both the mean along with the uncertainty/variance value needs to be obtained. As a test set predict for day numbers: 10, 30, 60, 90, 120, 150, 180, 210, 240, 250. Plot the training data points, (x axis-day, y-axis - daily vaccination), and the predicted mean and actual value at the test data points (with a different color). You may also draw vertical lines of size (+- variance) at the test points to visualize the uncertainty.

You may choose any GP prior, and corresponding hyperparameter values as you wish. However, choose a prior which captures the basic nature of the data. Assume a Gaussian noise with zero mean and variance of 0.1 denoting the error in reporting of vaccinations.

Bonus: Perform hyperparameter optimization/model averaging.

Submission: You need to submit two files - (i) a single program file (with your name, roll and dependencies mentioned in the header),

(ii) a write-up (single pdf file) with your name/roll and the details of method used, prediction accuracy, and visualization of results. For the program any language python/Julia/C/C++/java may be used. You may use basic data processing, linear algebra, random number generation libraries. But do not use GP functions directly.

Rules: Submission has to be individual. Plagiarism check will be performed and more than 70% overlap will be penalised.

You can use libraries, but not directly a GP implementation in a library.