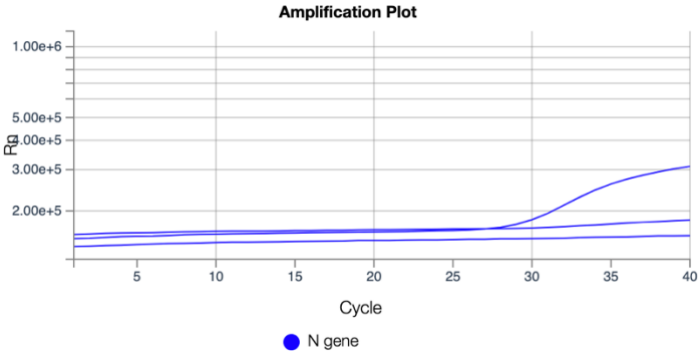
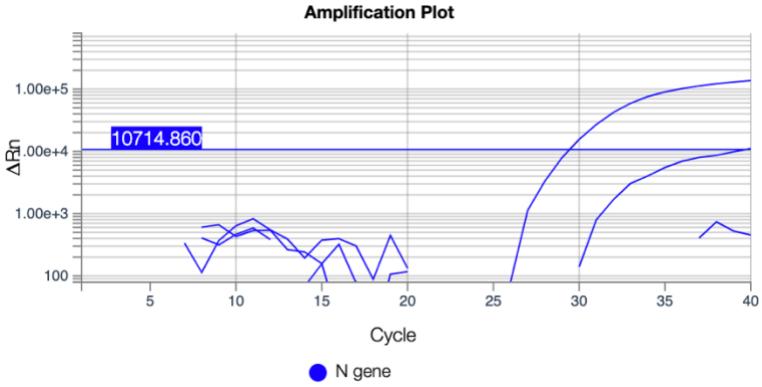
Huong Vu

Project Motivation

The project I am planning to work on is to improve COVID-19 RT-PCR test results using machine learning. The real-time reverse transcriptase polymerase chain reaction (RT-PCR) test is used widely as a gold standard reference test. However, current technology used in RT-PCR can only amplified DNAs up to 40 cycles and this limitation hinders the accuracy of the test especially when patients are in incubation period of COVID-19 with low amount of SARS-Cov-2 virus. If we could run PCR for longer cycles, we could have a more accurate result. The end goal of the big project is to predict positivity of a sample. Besides the direct impact of reducing false negative rate in COVID-19 PCR test results on public health, a big motivation for our project comes from the wide range of applications of RT-PCR in medical field such as detection of expressed genes or examination of transcript variants. Our result with COVID-19 data can be transferred and extended in other applications of RT-PCR.

Data





The left figure is a plot of raw data of a particular SARs gene from 3 different samples and the right figure is the processed data. By the mechanism of RT-PCR test, raw data of a COVID positive sample will look like a sigmoid curve after 40 cycles and a COVID negative sample data will have a linear trend. Lab technicians will set a threshold on processed data to determine positivity of the sample. If the sample passes the threshold within a range of cycles, the sample is considered as positive; otherwise, it is negative. Here, the middle curve is inconclusive on whether it should be positive or negative.

Plan

The goal of this project is to develop a method that fits the Rn curves (raw data). We would like to fit the following function which can cover both linear and sigmoid curves

$$y = s + \frac{a(x-t)}{(r + b|x-t|^{k})^{1/k}}$$

where a, b, r, k, s, and t are parameters that we would like to find using optimization.

Currently, there is an R package (qPCR) which implements the fitting of sigmoid functions of the form

$$y = c + \frac{d - c}{\left(1 + e^{(b(x-e))}\right)^{f}}$$

We will first investigate and evaluate this package and use its performance as the baseline to compare with our method. Right now, our plan is to try the zero-order and first-order methods to find a, b, r, k, s, and t parameters given a Rn curve.