APPLICATION OF MACHINE LEARNING METHODS TO COVID-19 EPIDEMIC PROCESS SIMULATION

Alireza Mohammadi, student of group 365
National Aerospace University “Kharkiv Aviation Institute”

Simulation and forecasting of the epidemic have been done based on various forecasting techniques and different data sources. These techniques are categorized into four main types:
1) Big data.
2) Social media/other communication media data.
3)Stochastic theory/mathematical models.
4) Data science/machine learning techniques.

Nowadays machine learning techniques are used worldwide for predictions due to their accuracy. Researchers have done predictions based on datasets that are available and used the best ML model as per the dataset. Various machine learning models have been implemented on coronavirus pandemic datasets of countries around the globe so far by different researchers. Generally, these models are categorized into two types of intelligent and statistical.

Intelligent models including Multilayer Perceptron, Hidden Markov Model, Hierarchical Bayes Model, Long Short-Term Memory (LSTM), etc. give a good trend for forecasting based on available datasets with a high rate of precision, they are complicated to implement and require specific well-structured datasets, though.

This research is based on Linear Regression and SIR models to analyze the available datasets and present a practical understanding of the trend, which as a result will lead to the ability to forecast the cases in upcoming days.

Statistical models including Regression and its sub models such as Linear Regression, ARIMA (Auto-Regressive Integrated Moving Average), VAR (Vector Auto Regression), SVR (Support Vector Regression), etc. are typically easier to implement than other models. They are capable to give a comprehensive view of the trend but not fully covering specific cases and major changes caused by an external factor.

Another statistical model which a famous epidemiologic model and is modest to implement nevertheless more efficient in considering external stimuli and sudden changes towards forecasting the trend is SIR (Susceptible, Infected, Recovered) and its sub models which present the trend with a higher rate of precision in forecasting future cases of COVID-19.

Regression analysis is a form of predictive modeling technique that investigates the relationship between dependent and independent variables. In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables).

Given a data set of n statistical units, a linear regression model assumes that the relationship between the dependent variable y and y and the p-vector of regressors x is linear. This relationship is modeled through a disturbance term or error variable ϵ - an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors.

Based on our model if the current situation with all rules and restrictions follows a steady state, we can expect the end of death cases in next year. Although the model doesn’t include unanticipated events such as inventing COVID19 vaccine, loosen or strengthen the restrictions, the influence of changing seasons and etc., This model helps us to have a comprehensive understanding of when we can expect the end of pandemic for 2021.

SIR is a simple model that considers a population that belongs to one of the following states:

1) Susceptible (S). The individual hasn't contracted the disease, but she can be infected due to transmission from infected people.
2) Infected (I). This person has contracted the disease.
3) Recovered/Deceased (R).

The disease may lead to one of two destinies: either the person survives, hence developing immunity to the disease or the person is deceased.

The SIR model and its extended modifications, a mathematical model in various forms have been used in previous studies to model the spread of COVID-19 within communities. The SIR model we introduce here is given by the same simple system of three ordinary differential equations (ODEs) as the classic SIR model and can be used to gain a better understanding of how the virus spreads within a community of variable populations in time when surges occur. Importantly, it can be used to make predictions of the number of infections and deaths that may occur in the future and provide an estimate of the time scale for the duration of the virus within a community. It also provides us with insights on how we might lessen the impact of the virus, which is nearly impossible to discern from the recorded data alone. Consequently, our SIR model can provide a theoretical framework and predictions that can be used by government authorities to control the spread of COVID-19.

*Scientific Advisor – Chumachenko D., Ph.D., Associate Professor of Mathematical modelling and artificial intelligence department.