My idea of project for this course is about COVID Global Forecasting.

In the context of the global COVID-19 pandemic, Kaggle has launched several challenges in order to provide useful insights that may answer some of the open scientific questions about the virus. This is the case of the COVID19 Global Forecasting, in which participants are encouraged to fit worldwide data in order to predict the pandemic evolution, hopefully helping to determine which factors impact the transmission behavior of COVID-19.

In this course, we will learn some kinds of machine learning algorithms for recognition and prediction task. Therefore, I attempt to use the machine learning, such as regression, to do the COVID Global Forecasting task mentioned above. It is an interesting and timely idea to think about how to use ML algorithm to accomplish this task.

Essentially, COVID Global Forecasting is a kind of prediction task.

Kaggle provides us the raw data about the information of COVID data in several countries.

Patrick’s model (SIR regression model) will be used as my baseline to develop my own model.

My plan for this project is as follows:

First, preprocess the raw data, such as clear the missing data.

Second, reproduce the aforementioned baseline model.

Third, develop my own model to accomplish the forecasting task.

Fourth, train my own model to obtain the regression model’s parameters.

Fifth, do comparative experiments to illustrate that our model outperforms the baseline we choose.