PH244
Big Data: A Public Health Perspective
Project IV: Understanding COVID19

Amid the current pandemic, it is an international imperative to better understand COVID19. Kaggle is hosting multiple data challenges. We are going to investigate two of them. The data is actually "small" compared to the ones we have been working on. But I think it'd be great to use the knowledge we talk about in this Big Data class to help us all improve our understanding of this disease.

Data:

(a) COVID19 Global Forecasting:
Link: `https://www.kaggle.com/c/covid19-global-forecasting-week-3/`.
A copy of data can be found on bCourses, under `/Files/Project/Project-IV/`.

(b) UNCOVER COVID-19 Challenge:
Link: `https://www.kaggle.com/roche-data-science-coalition/uncover/`.
A copy of data can be found on bCourses, under `/Files/Project/Project-IV/`.

Problems:

(1) Forecast the number of daily cases and deaths: You may choose a cutoff date, and use the numbers before that date as the training samples, and the numbers after that date as the testing samples. Alternatively, you may choose a subset of countries as the training samples, and the rest of countries as the testing samples.

(2) There are multiple tasks in this competition. Choose ***only one*** task from (b).

Suggestions:

(i) You can be creative here, as long as the task is something interesting to investigate, and can help improve our understanding of the disease. For instance, it can be an interactive map, or some comparison between different countries or states. It can even be some modification of one of the tasks in (b).

(ii) You are welcome to see what other people on Kaggle are doing for different tasks. If you do so, please ***cite*** other people's work properly. It would be best if you can introduce something of your own, even though it may be only a small modification and/or improvement.

(iii) You can choose to work by yourself, or team up with another classmate. Each group should be no more than 2 students.