

Analyses of daily COVID-19 cases across nations

Group11: Sibe Liu, Xue Jin, Yuchen Qi, Xinru Wang

05/01/2020

Introduction

COVID-19

Since its first outbreak in January, the novel coronavirus (COVID-19) has been spreading rapidly through China and expanded to touch nearly every corner of the globe. Hundreds of thousands of people around the world have been sickened and over 200,000 have died. Efforts to contain the spread of the Covid-19 pandemic are now the top priority of governments. To make scientific decisions, such as quarantine, active monitoring, border controls, and lockdown, it is particularly crucial for policymakers to accurately predict how the spread of COVID-19 will change over time.

A logistic growth curve can be an effective way to capture the trajectory of cumulative cases of COVID-19. Characterized by an S-shaped curve, logistic growth model is approximately exponential at first, and growth rate accelerates as it approaches the midpoint of the curve but begins to decelerate as it approaches the model's upper bound, called the carrying capacity. In the COVID-19 case, this maximum limit would be the maximum number of cases a region can reach. The more people who have the virus, the more rapidly it spreads, and the growth will necessarily diminish when everybody is sick, which make the logistic model a good one to study the spread.

Objectives

To help predict future spread of Covid-19 and to identify risk factors, our project aims to fit a logistic curve to the cumulative confirmed COVID-19 cases in each region of the world by developing an optimization algorithm and implement K-mean and Gaussian mixture model (with EM algorithm) to cluster these curves based on the fitted parameters.

Dataset

The dataset is a subset of the open data, which contains the cumulative number of confirmed cases and death of COVID-19 from Jan 21 to March 23 from 163 countries/regions. Eight variables are recorded as following:

- Id: Record ID
- Province/State: The local state/province of the record;
- Country/Region: The country/region of the record;
- Lat: Latitude of the record;
- Long: Longitude of the record;
- Date: Date of the record;
- ConfirmedCases: The number of confirmed case on that day;
- Fatalities: The number of death on that day;

Statistical Methods

Adam Algorithm

EM Algorithm

Result

Discussion

Task 1:

Task 2

Conclusions

Figures