

CSE506
Introduction to Data Mining
Assignment 1

Submitted By-
Gaurav Lodhi (MT19063)
NitindeepSingh (MT19069)

Methodology:

We have read the json file converting it into dataframe as well as python dictionaries in different questions as per the need. Individual modular functions have been created for each question as per the code sample provided (Sample.py). Whenever any function is called, first the date is validated if the range is valid or not, if the range is valid, then the date yyyy-mm-dd is converted into the format present in the json dataset else the specific error is displayed on the screen. All the variables are returned as per the instructions.

Assumptions:

- Cumulative count is taken in Q1_1 and Q2_1.
- Q3 is done in two ways:
 - **1. Gradient Descent Method :-** To improve the scalability of the algorithm in case of increasing no of attributes. It is computationally cheaper when there is multivariate data.
 - **2. Normal Equation Method:-** Here weights are directly calculated to fit the regression line by using the following mathematical formula.

$$W=(X^T X)^{-1} (X^T Y)$$

It is easy to use but it takes high computation because it involves matrix inverse and multiplication operations.

Following are the results of the Assignment 1:

MT19069
MT19063

Q1_1

confirmed_count: 4110211
recovered_count: 3177666
deceased_count: 70094

Q1_2

confirmed_count: 188193
recovered_count: 163785
deceased_count: 4538

Q1_3

confirmed_count: 1072055
recovered_count: 800359
deceased_count: 30813

Q1_4

Confirmed

Lowest affected State is: ['mh']

Lowest affected State count is: 883862

Recovered

Lowest affected State is: ['mh']

Lowest affected State count is: 636574

Deceased

Lowest affected State is: ['mh']

Lowest affected State count is: 26275

Q1_5

Confirmed

Lowest affected State is: ['dd', 'ld', 'un']

Lowest affected State count is: 0

Recovered

Lowest affected State is: ['dd', 'ld', 'un']

Lowest affected State count is: 0

Deceased

Lowest affected State is: ['dd', 'ld', 'mz', 'un']

Lowest affected State count is: 0

Q1_6

Confirmed

Day: 23-Jun-20

Count: 3947

Recovered

Day: 20-Jun-20

Count: 7725

Deceased

Day: 16-Jun-20

Count: 437

Q1_7

State: an

Active Cases: 343

State: ap

Active Cases: 100880

State: ar

Active Cases: 1525

State: as

Active Cases: 28404

State: br

Active Cases: 16735

State: ch

Active Cases: 2143

State: ct

Active Cases: 22320

State: dd

Active Cases: 0

State: dl

Active Cases: 19870

State: dn

Active Cases: 301

State: ga

Active Cases: 4945

State: gj

Active Cases: 16266

State: hp

Active Cases: 2023

State: hr

Active Cases: 14912

State: jh

Active Cases: 14980

State: jk

Active Cases: 9547

State: ka

Active Cases: 100224

State: kl

Active Cases: 21867

State: la

Active Cases: 834

State: ld
Active Cases: 0

State: mh
Active Cases: 221013

State: ml
Active Cases: 1374

State: mn
Active Cases: 1872

State: mp
Active Cases: 15687

State: mz
Active Cases: 349

State: nl
Active Cases: 701

State: or
Active Cases: 25856

State: pb
Active Cases: 15870

State: py
Active Cases: 5163

State: rj
Active Cases: 14996

State: sk
Active Cases: 561

State: tg
Active Cases: 32405

State: tn
Active Cases: 51580

State: tr
Active Cases: 5905

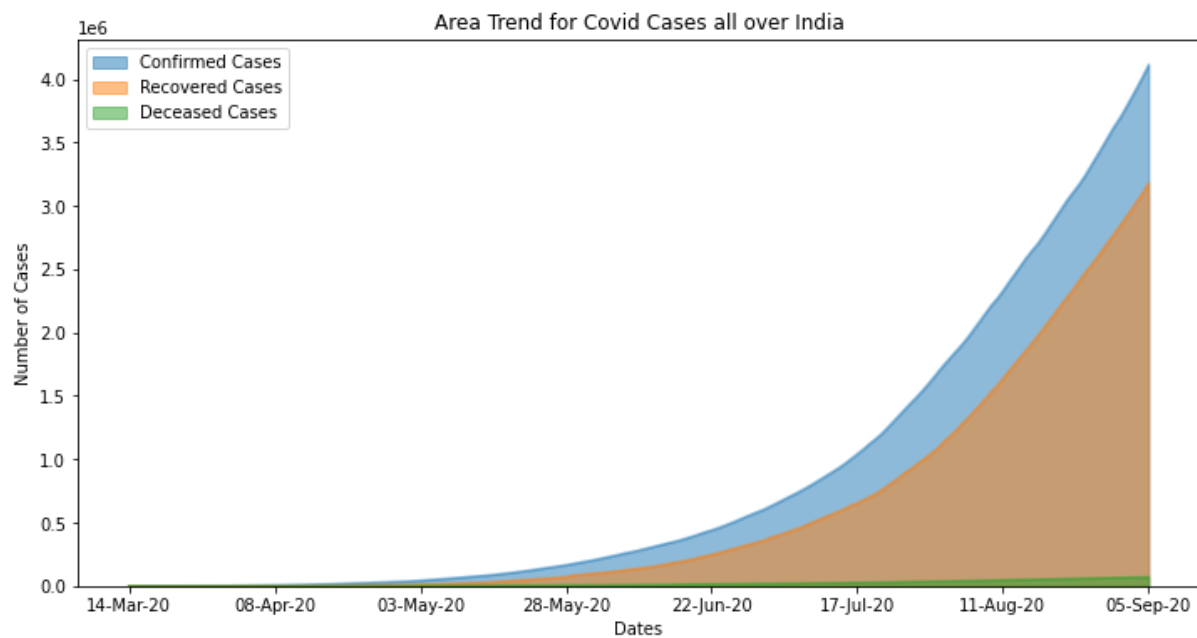
State: un
Active Cases: 0

State: up
Active Cases: 59963

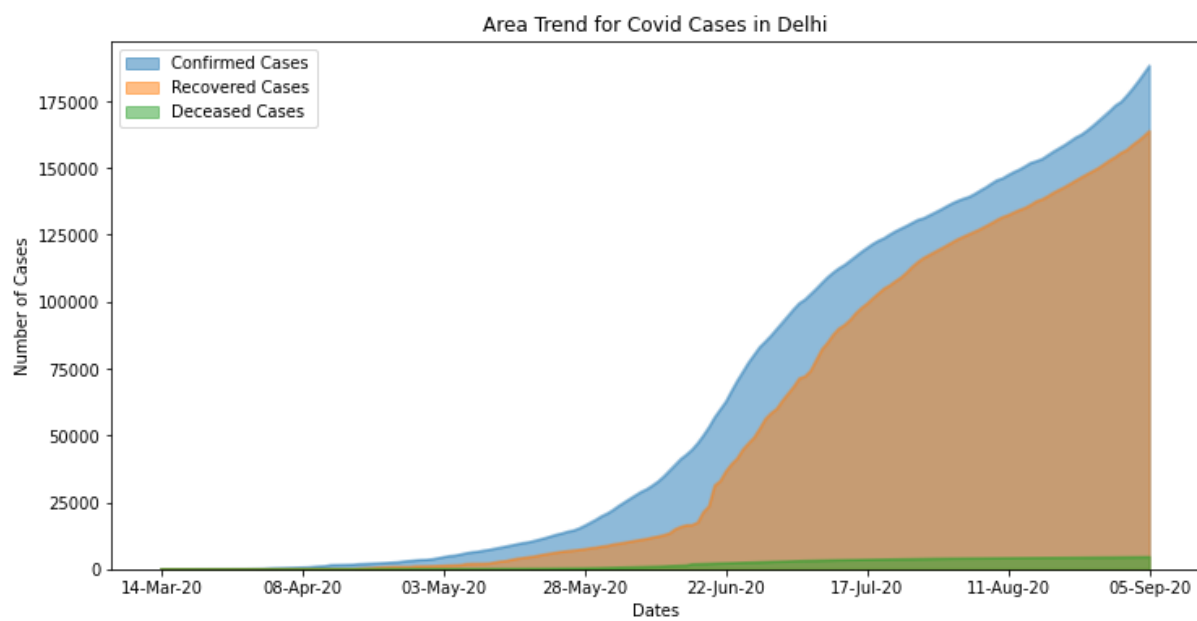
State: ut
Active Cases: 7649

State: wb
Active Cases: 23390

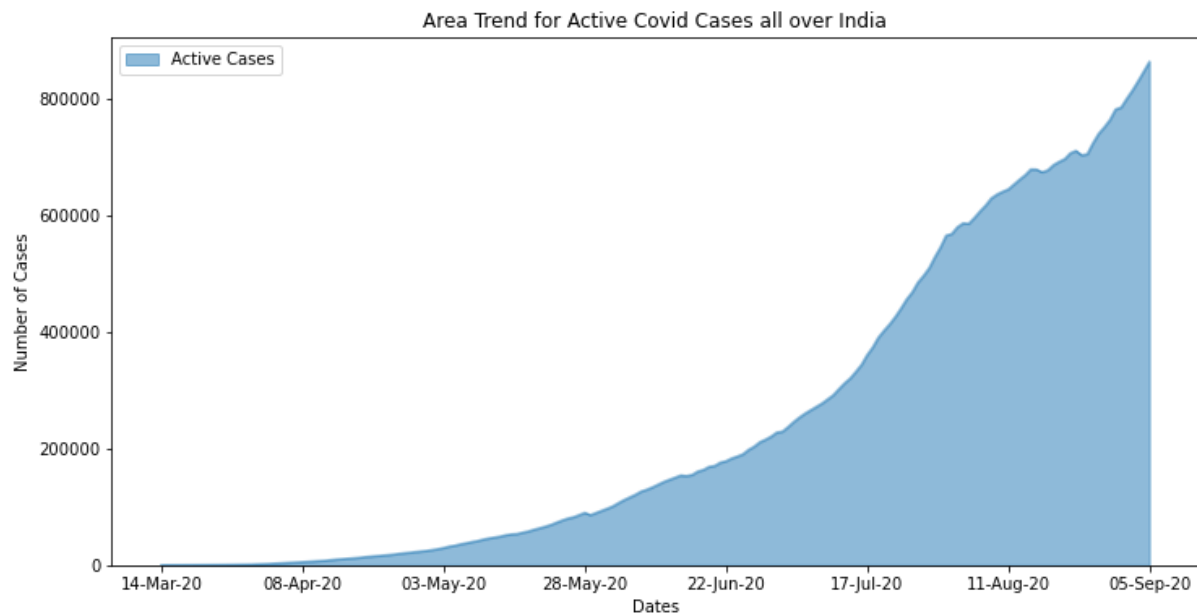
Q2_1



Q2_2



Q2_3

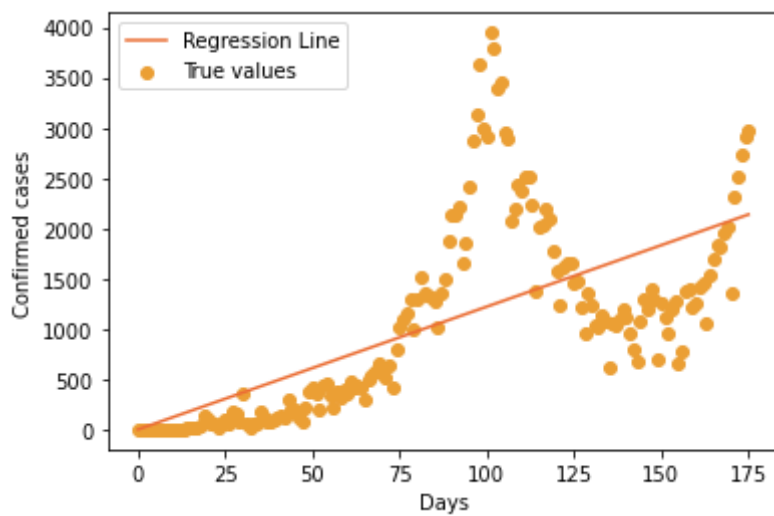


Q3

Normal Equation Method:

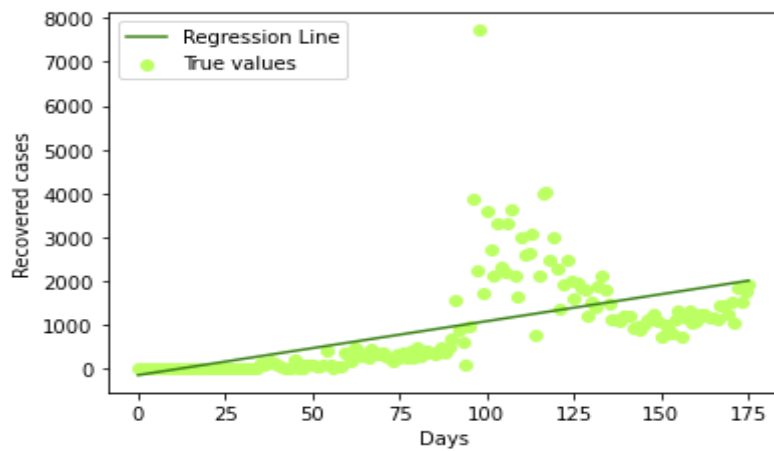
Slope of confirmed cases: [12.21426921]

Intercept of confirmed cases: [-11.68441558]



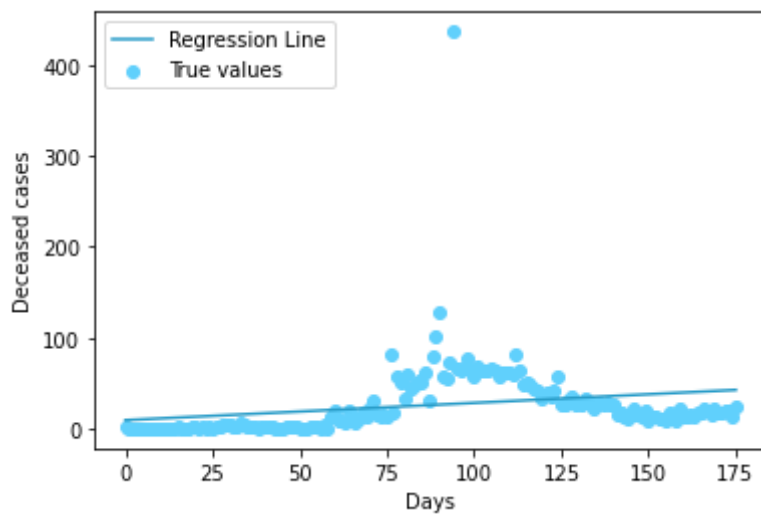
Slope of recovered cases: [12.30552829]

Intercept of recovered cases: [-158.44266234]



Slope of deceased cases: [0.19023333]

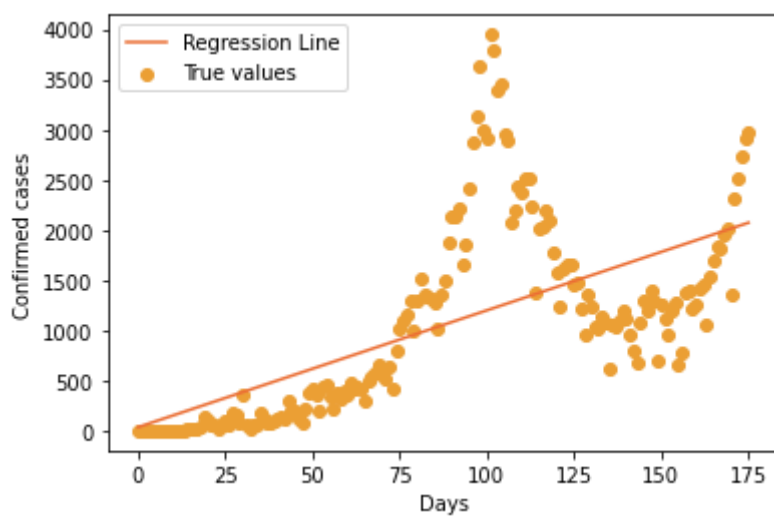
Intercept of deceased cases: [8.94844156]



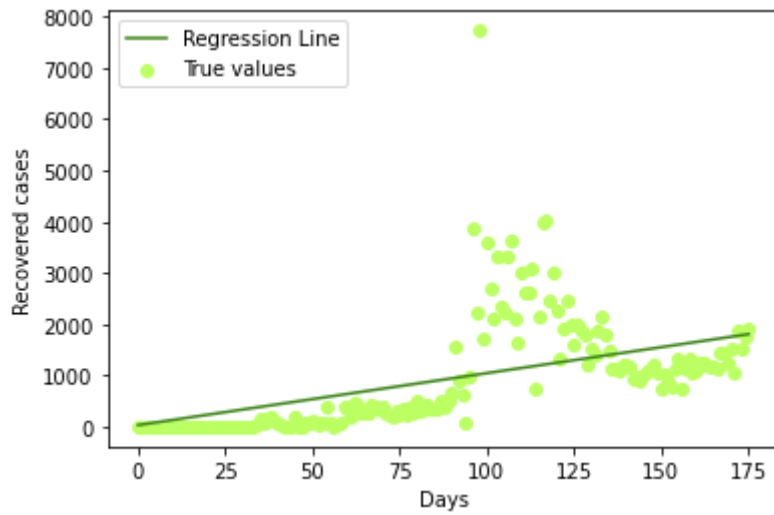
Using Gradient Descent Method:

Slope of confirmed cases: 11.640591492755062

Intercept of confirmed cases: 23.28118298551012



Slope of recovered cases: 10.130845874399617
Intercept of recovered cases: 20.26169174879925



Slope of deceased cases: 0.2806959036421252
Intercept of deceased cases: 0.5613918072842509

