

ANALISI DEL COVID-19

in Italia con gli Open Data

Speaker:

Vincenzo Lavorini,
Data Scientist Researcher

ONLINE

30 marzo - 18.30

Hosted by



In partnership with



MEETUP



Disclaimer

Non è uno studio epidemiologico!

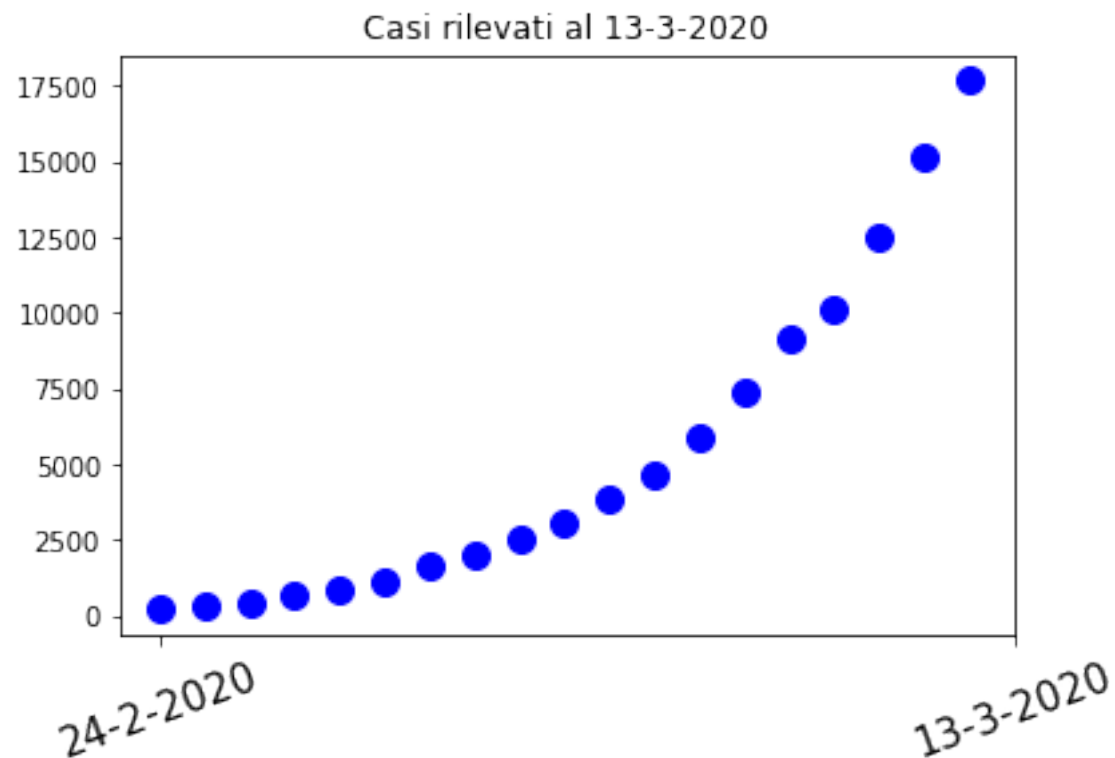
Obiettivo dello studio

- Valutare approccio da non esperti in epidemiologia
- *“tutti i modelli sono sbagliati, ma qualcuno di essi è utile”*
- Iterazioni:
 - Ideiamo modello
 - Valutiamo
 - Miglioriamo
- Obiettivo: stimare numero di casi totali e fine dell'ondata

Open Data

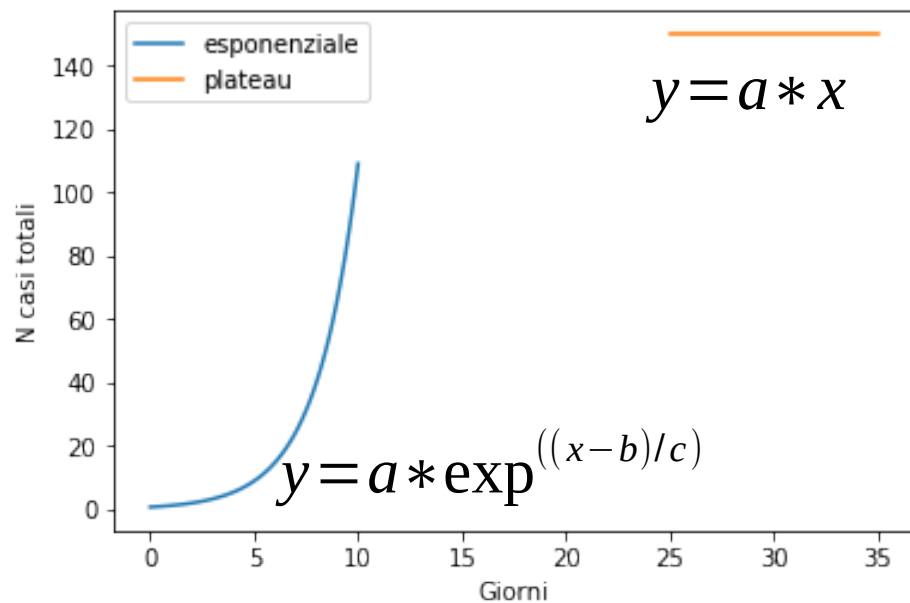
<https://github.com/pcm-dpc/COVID-19>

Totale casi,
Giorno dopo giorno



Ipotesi

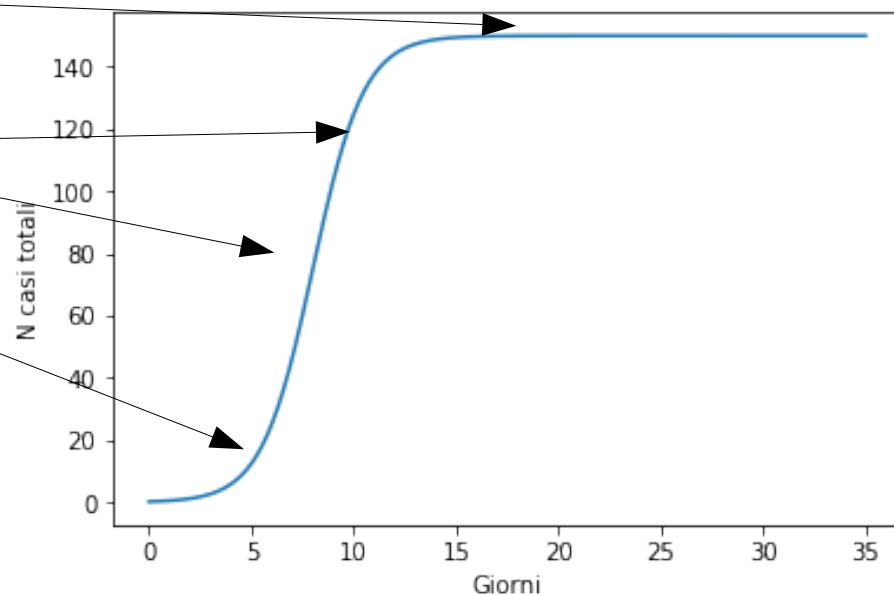
- In fase iniziale, crescita esponenziale del numero totale di casi
- A regime, nessun nuovo caso (plateau)
- Dobbiamo unirle:
 - Curva liscia, no punti singolari



Funzione sigmoide

- Scelta forzatamente semplice
- Definita da tre parametri :
 - Plateau
 - Flesso
 - Crescita

$$y = \frac{\text{plateau}}{1 + e^{\frac{-(x - \text{flesso})}{\text{crescita}}}}$$



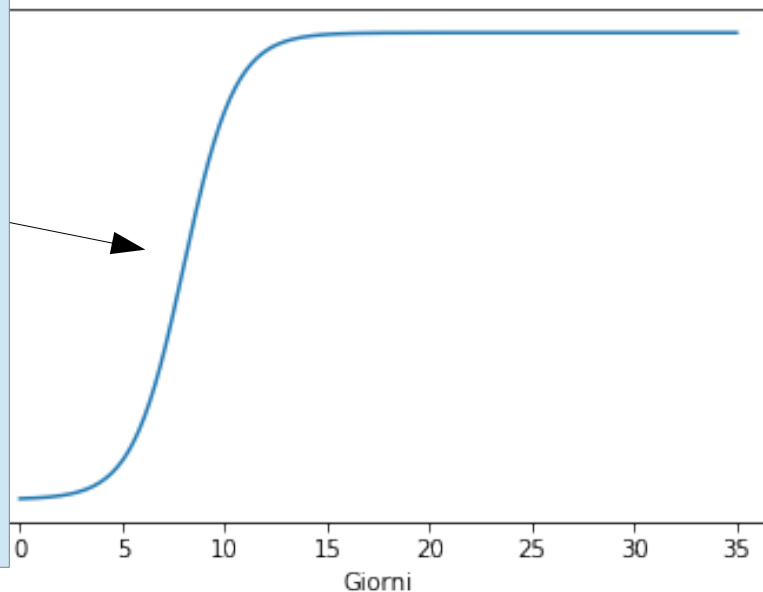
Funzione sigmoide

Costo fortemente convesso

Se non raggiungiamo il punto di flesso :

- La crescita rimane esponenziale
→ non possiamo predire il plateau

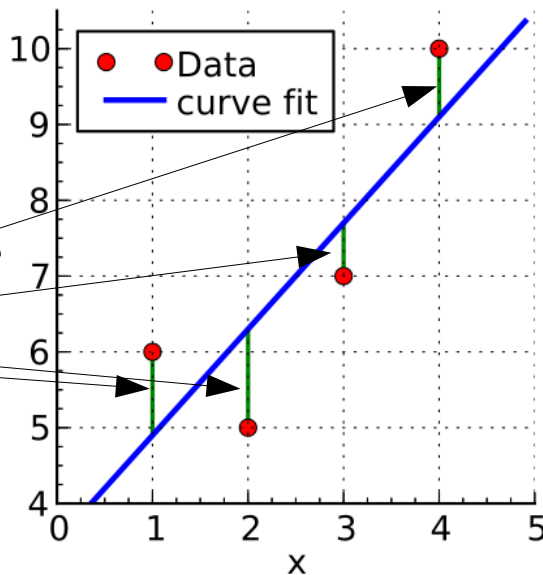
$$y = \frac{\text{plateau}}{1 + e^{\frac{-x - \text{flesso}}{\text{pendenza}}}}$$



Fitting : ottimizzazione parametri

- Quali valori dare ai parametri della curva per descrivere i dati ?
- Operazione di Supervised Machine Learning
- Funzione da minimizzare :
 - Somma dei quadrati dei residui (L2)

$$L = \sum_0^i r_i^2$$

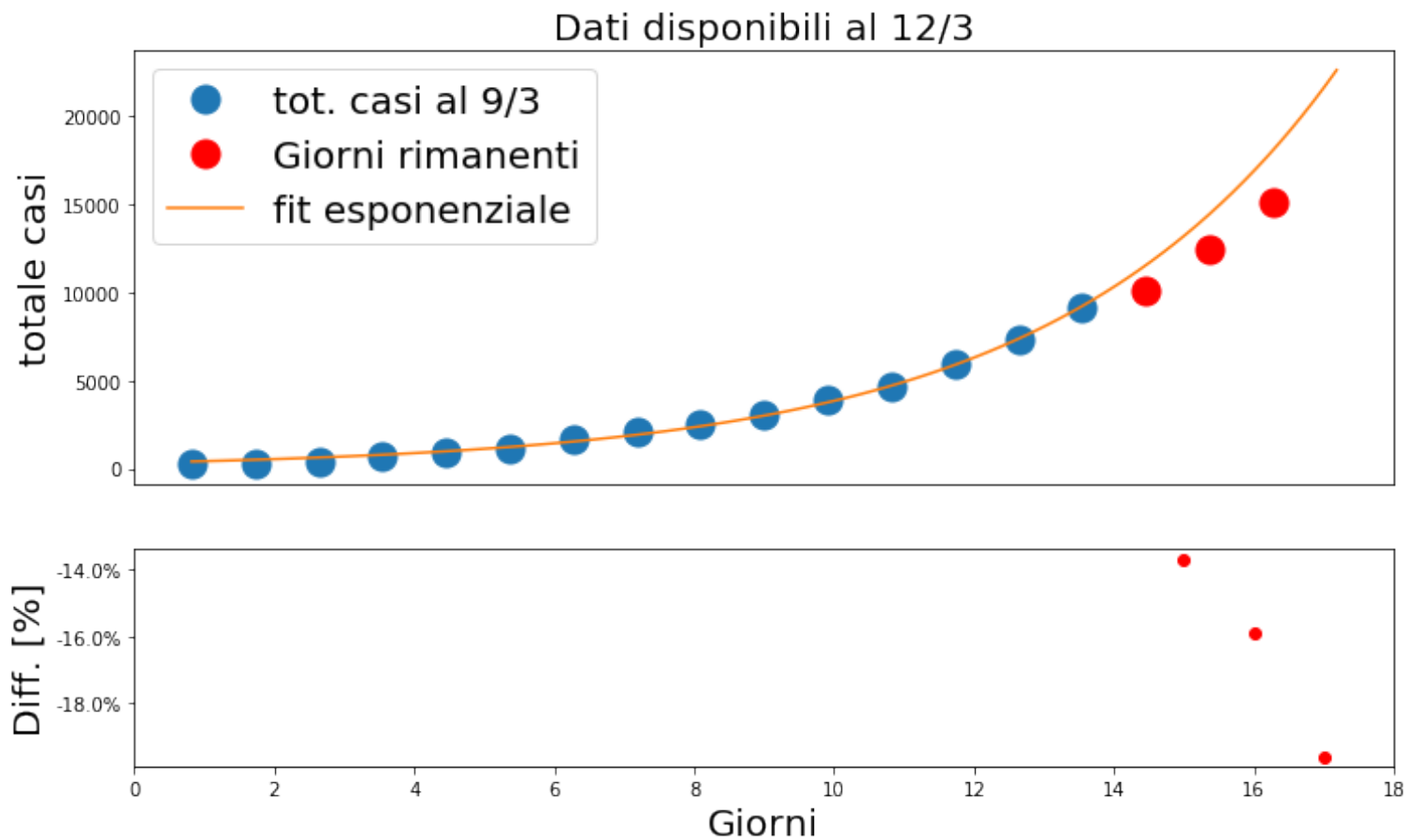


Cerchiamo punto di flesso

Per trovarlo analizziamo il n. tot di casi giorno per giorno:

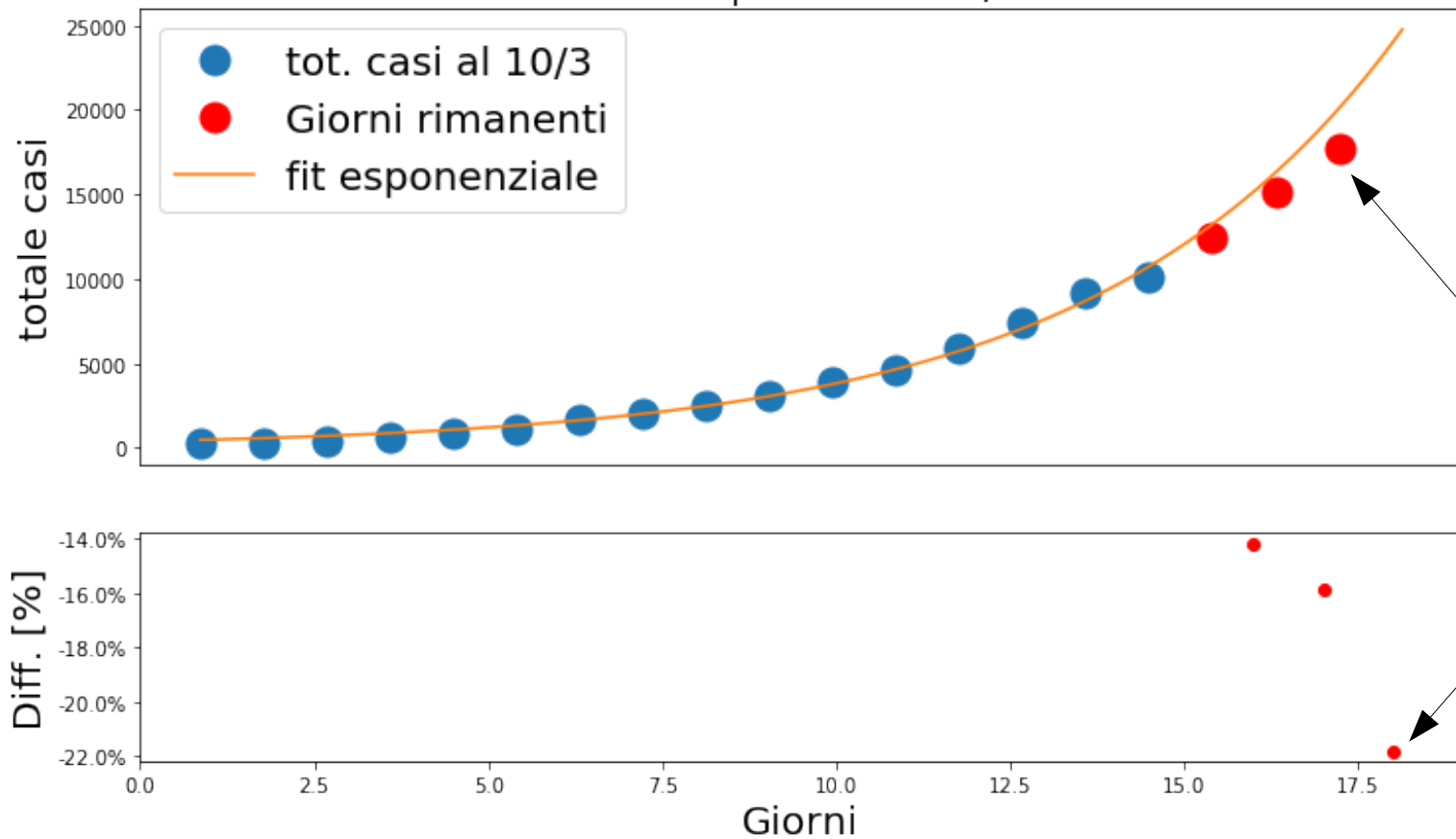
- Adattiamo curva esponenziale ai dati al giorno G-3
- Se i tre giorni rimanenti cadono al di sotto dell'esponenziale :
 - **Abbiamo passato il flesso !**

Situazione al 12 Marzo



Situazione al 13 Marzo

Dati disponibili al 13/3



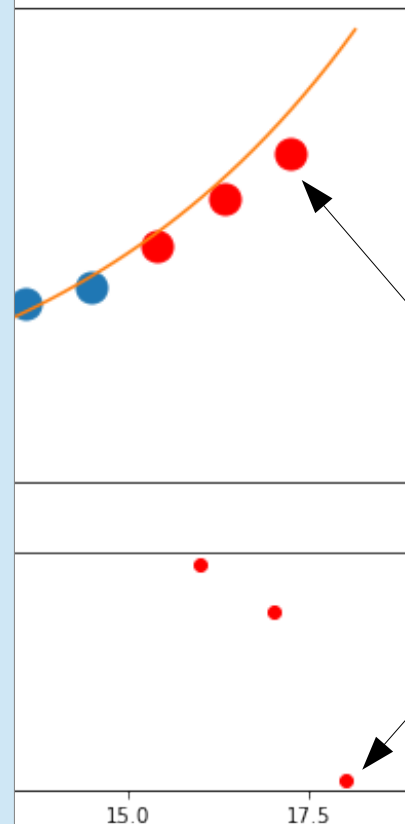
!!!

Situazione al 13 Marzo

Dati disponibili al 13/3

Inoltre :
Lockdown zone rosse del 7 Marzo
Periodo medio incubazione COVID-19 : 5 giorni

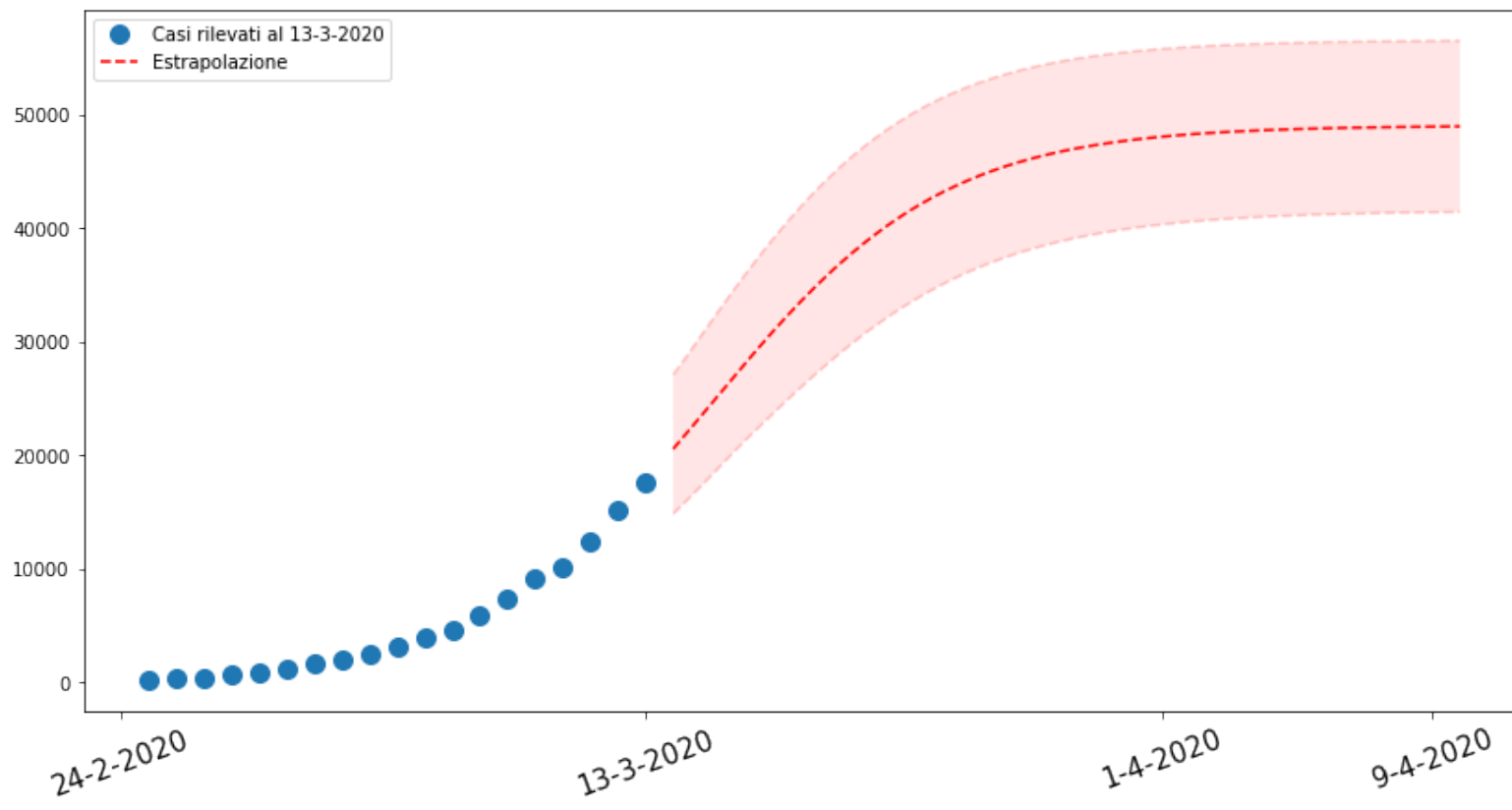
...flesso raggiunto ! (FALSO)



!!!

Previsione (errata)

Totale casi attesi in data 9-4-2020: tra 41000 e 56000
98% del totale in data 1-4-2020



Previsione (errata)

Tecniche meno sensibili all'emotività?

e 56000



2020

9-4-2020

Programmazione probabilistica

Metodi basati su **Statistica Bayesiana** e **Markov Chain Monte Carlo (MCMC)**

Statistica Classica (Frequentista)

Presuppone $N \rightarrow \infty$

Statistica Bayesiana

considera anche N piccoli

Programmazione probabilistica

Metodi basati su **Statistica Bayesiana** e **Markov Chain Monte Carlo (MCMC)**

Statistica Classica (Frequentista)

Presuppone $N \rightarrow \infty$

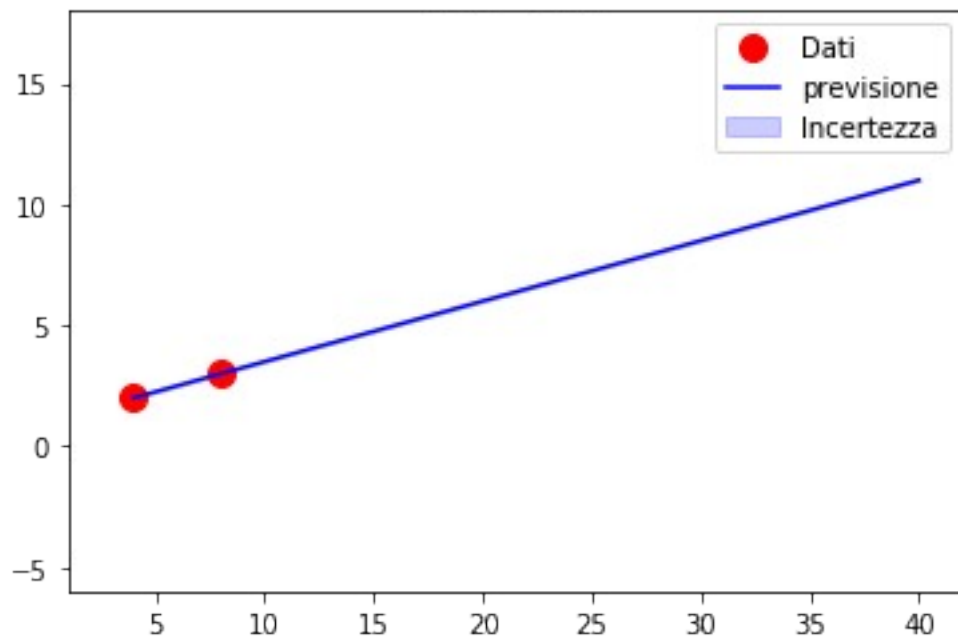
Statistica Bayesiana

In generale, no soluzioni analitiche
Quindi: **MCMC**

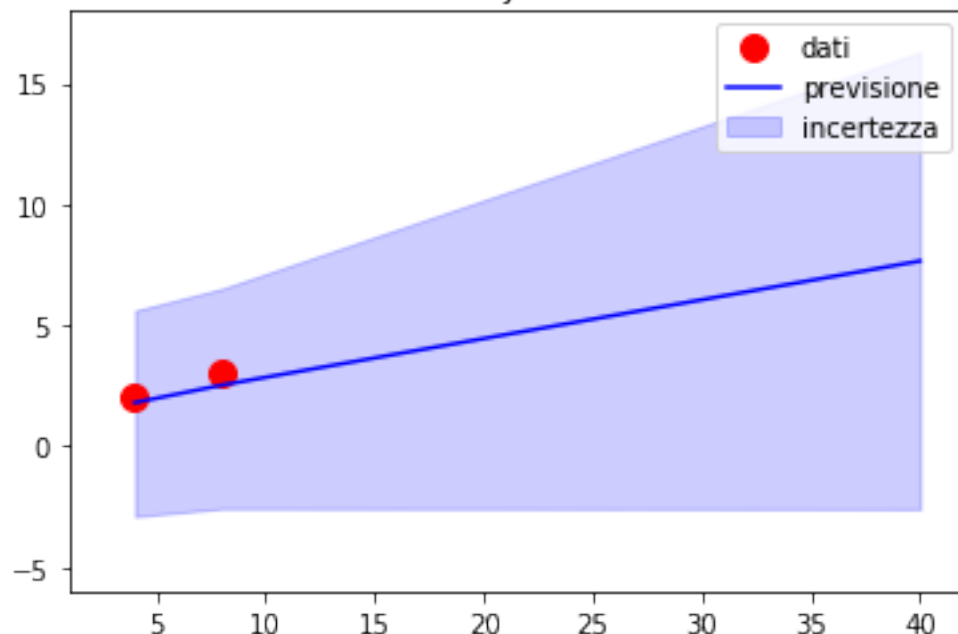
Confronto: fit lineare 2 punti

Consideriamo (4, 2) ed (8, 3). Adattiamo funzione lineare ($y=a+b*x$), e valutiamo la funzione ad $x=40$

Fit classico

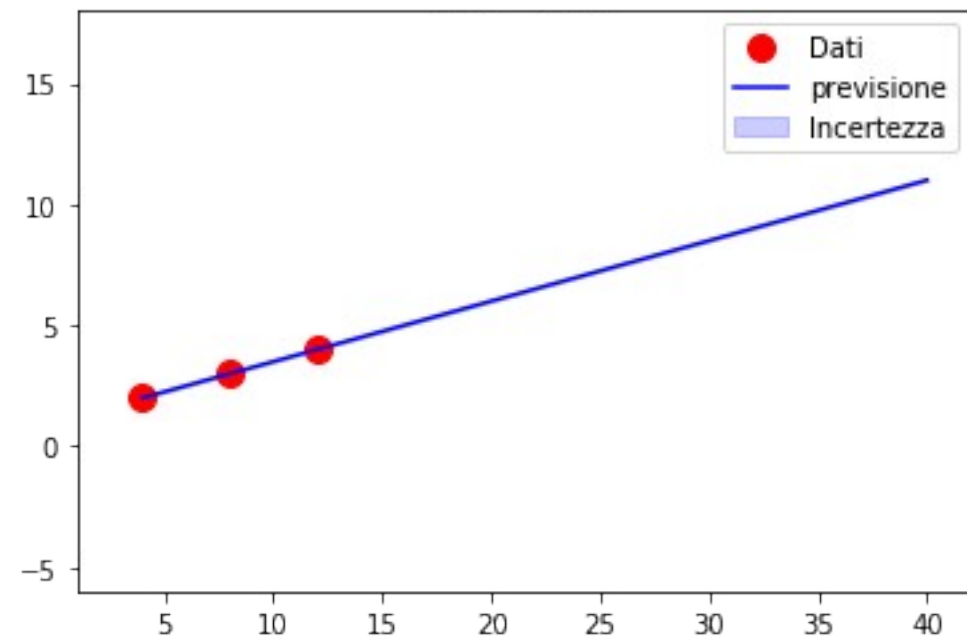


Fit Bayesiano

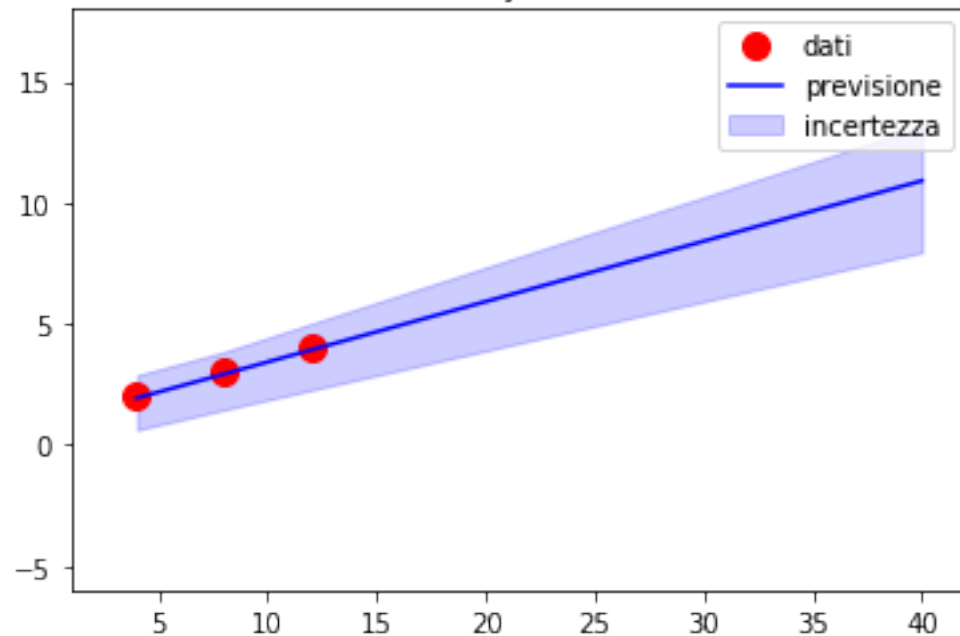


Confronto: fit lineare 3 punti

Fit classico



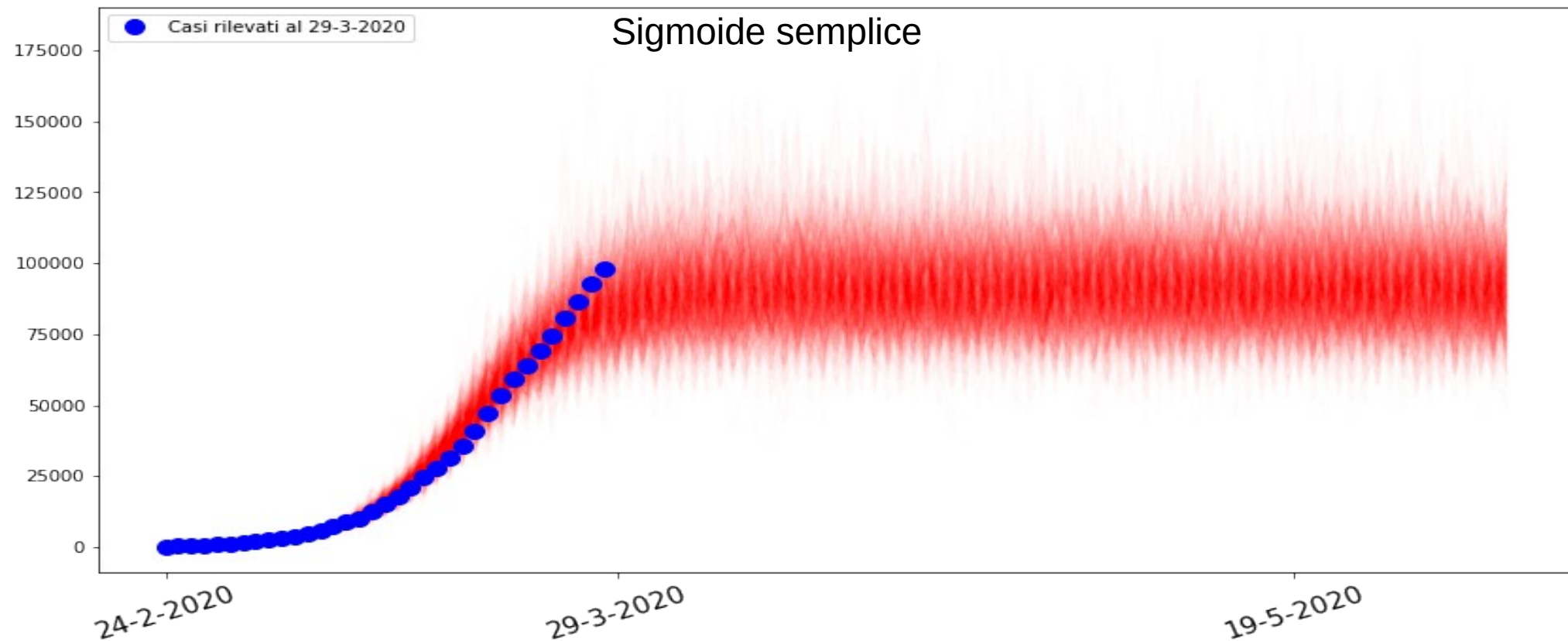
Fit Bayesiano



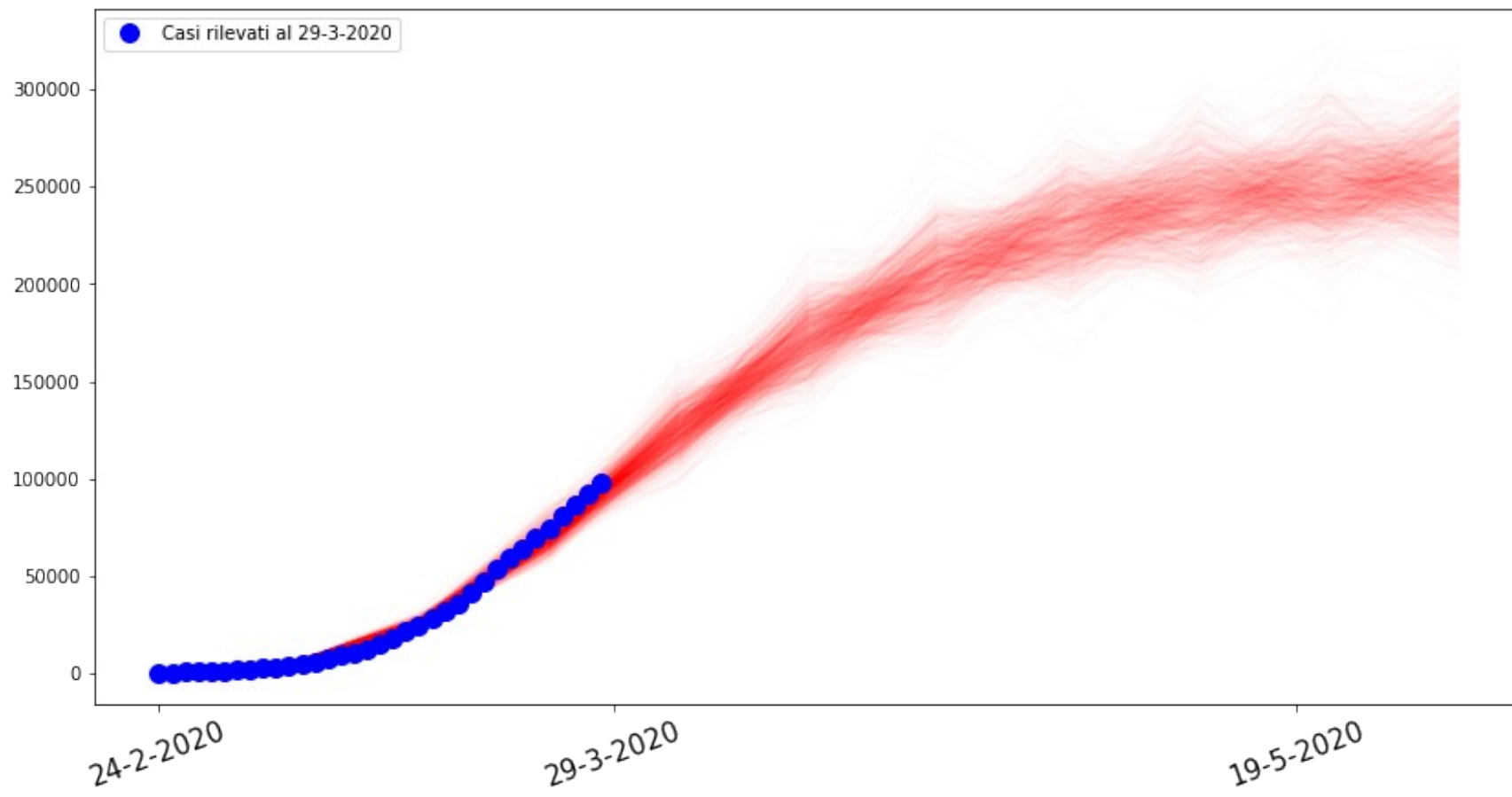
Fit bayesiano al 13 Marzo

Vedi notebook

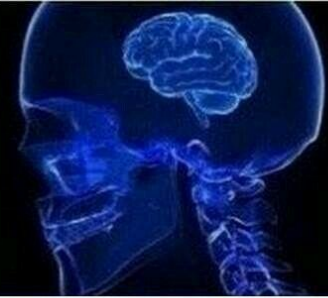
Con dati di ieri (29 Marzo)



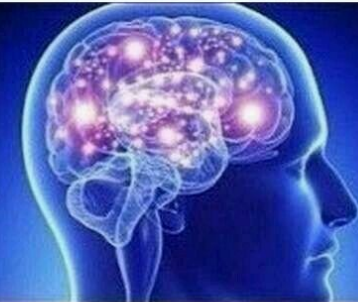
Nuovo modello: Gompertz



Fit semplice



Fit Bayesiano

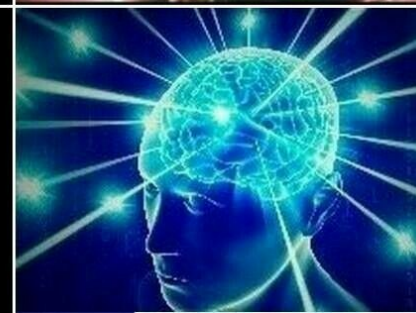


SIR



SIR con Fit bayesiano

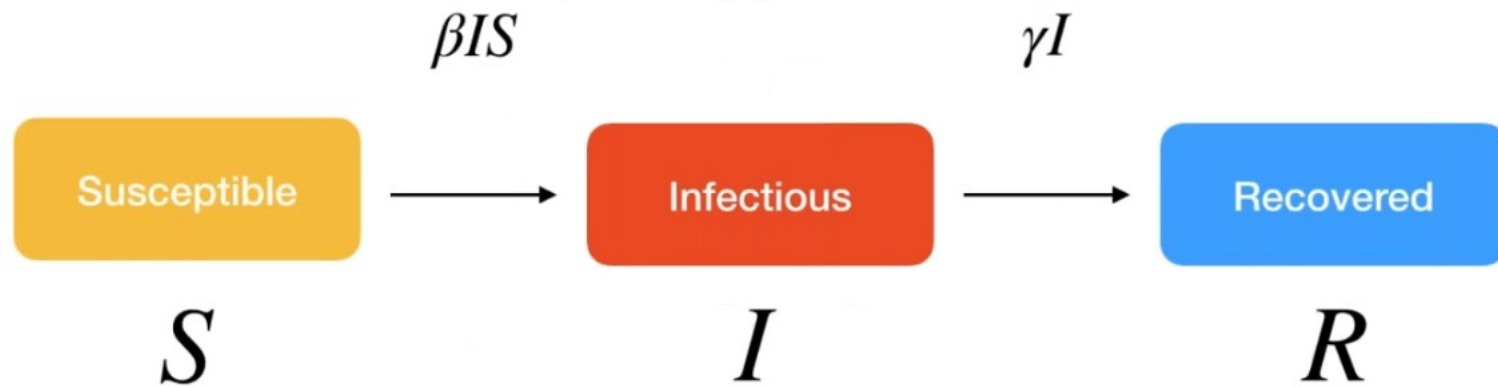
https://github.com/twiecki/covid19/blob/master/covid19_sir.ipynb

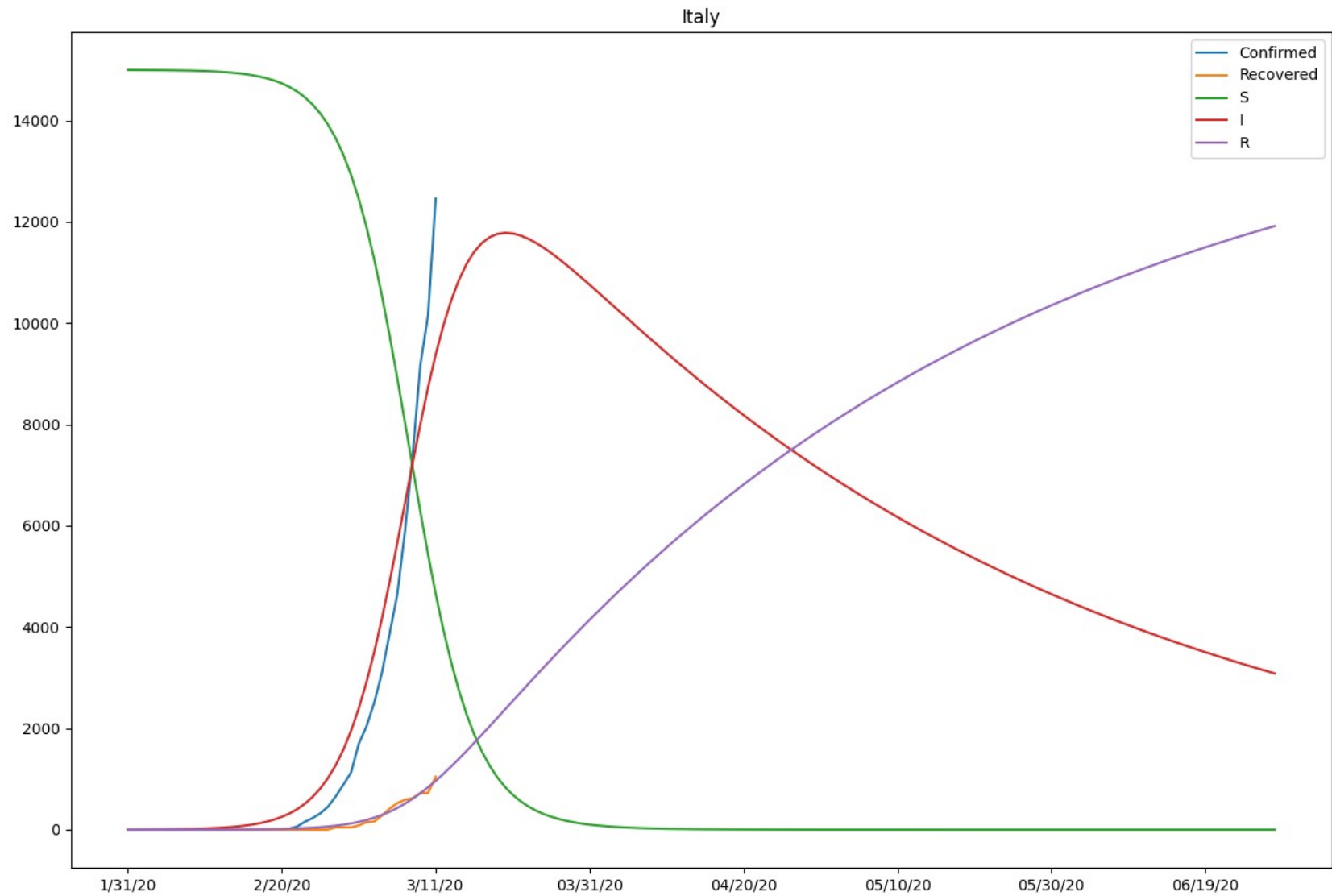


Conclusioni

- Si possono analizzare dati senza avere conoscenza specifica della materia.
 - Non fidarsi troppo delle proprie posizioni
 - Non sostituirsi ad esperti in materia

Backup: SIR





<https://www.lewuathe.com/covid-19-dynamics-with-sir-model.html>